BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 626 998 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
31.10.2001 Bulletin 2001/44

(21) Application number: **93906130.5**

(22) Date of filing: **18.02.1993**

(51) Int Cl.[7]: **C12N 5/10**, A01N 63/02

(86) International application number:
PCT/US93/01544

(87) International publication number:
WO 93/17098 (02.09.1993 Gazette 1993/21)

(54) **PRODUCTION OF VIRAL RESISTANT PLANTS VIA INTRODUCTION OF UNTRANSLATABLE PLUS SENSE VIRAL RNA**

PRODUKTION VIRUS-RESISTENTER PFLANZEN DURCH EINFÜHRUNG VON NICHT-TRANSLATIERBARER VIRALER PLUS-STRANG-RNA

PRODUCTION DE PLANTES RESISTANT A DES VIRUS PAR INTRODUCTION D'ARN VIRAL DE NON TRANSLATION A SENS POSITIF

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **19.02.1992 US 838509**

(43) Date of publication of application:
07.12.1994 Bulletin 1994/49

(73) Proprietor: **THE STATE OF OREGON ACTING BY AND THROUGH THE OREGON STATEBOARD OF HIGHER EDUCATION ON BEHALF OF THE UNIVERSITY OF OREGON**
Corvallis, OR 97331-2140 (US)

(72) Inventors:
- **DOUGHERTY, William, G.**
  Philomath, OR 97370 (US)
- **LINDBO, John, A.**
  Corvallis, OR 97330 (US)

(74) Representative: **Gowshall, Jonathan Vallance**
FORRESTER & BOEHMERT
Pettenkoferstrasse 20-22
80336 München (DE)

(56) References cited:
EP-A- 0 242 016
EP-A- 0 298 918
EP-A- 0 325 066
EP-A- 0 425 004
EP-A- 0 426 195
EP-A- 0 504 869
WO-A-92/13089
WO-A-92/13090

- **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 87, August 1990 WASHINGTON US, pages 6291-6295, STANLEY, J., ET AL. 'Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants'**
- **VIROLOGY, vol. 164, 1988 pages 383-389, VAN DUN, C.M.P., ET AL. 'Transgenic tobacco expressing tobacco streak virus or mutated alfalfa mosaic virus coat protein does not cross-protect against alfalfa mosaic virus infection'**
- **ANNUAL MEETING OF THE AMERICAN PHYTOPATHOLOGICAL SOCIETY, ST. LOUIS, MISSOURI, USA, AUGUST 17-21, 1991. PHYTOPATHOLOGY 81 (10). 1991. 1174. , LINDBO J A ET AL 'TRANSGENIC NICOTIANA-TABACUM PLANTS EXPRESSING VARIOUS FORMS OF THE TOBACCO ETCH VIRUS TEV COAT PROTEIN ARE RESISTANT TO TEV INFECTION.'**
- **Molecular Plant-Microbe Interactions, Volume 5, No. 2, issued March 1992, LINDBO et al.: "Pathogen-Derived Resistance to a Potyvirus: Immune and Resistant Phenotypes in Transgenic Tobacco Expressing Altered Forms of a Potyvirus Coat Protein Nucleotide Sequence", pages 144-153, see entire document.**

EP 0 626 998 B1

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- Virology, Volume 189, No. 2, issued August 1992, LINDBO et al.: "Untranslatable Transcripts of the Tobacco Etch Virus Coat Protein Gene Sequence can Interfere with Tobacco Etch Virus Replication in Transgenic Plants and Protoplasts", pages 725-733, see entire document.
- Molecular Plant-Microbe Interactions, Volume 4, No. 3, issued May 1991, KAWCHUK et al.: "Sense and Antisense RNA-Mediated Resistance to Potato Leafroll Virus in Russet Burbank Potato Plants", pages 247-253, see entire document.
- Plant Molecular Biology, Volume 17, issued 1991, VAN DER WILK et al.: "Expression of the Potato Leafroll Luteovirus Coat Protein Gene in Transgenic Potato Plants Inhibits Viral Infection", pages 431-439, see entire document.
- Journal of General Virology, Volume 72, issued August 1991, MARSH et al.: "Artificial Defective Interfering RNAs Derived from Brome Mosaic Virus", pages 1787-1792, see entire document.
- Proceedings of the National Academy of Sciences USA, Volume 88, issued August 1991, DAY et al.: "Expression of an Antisense Viral Gene in Transgenic Tobacco Confers Resistance to the DNA Virus Tomato Golden Mosaic Virus", pages 6721-6725, see entire document.
- Virology, Volume 175, issued 1990, POWELL et al.: "Protection Against Tobacco Mosaic Virus Infection in Transgenic Plants Requires Accumulation of Coat Protein Rather than Coat Protein RNA Sequences", pages 124-130, see entire document.
- Virology, Volume 154, issued 1986, ALLISON et al.: "The Nucleotide Sequence of the Coding Region of Tobacco Etch Virus Genomic RNA: Evidence for the Synthesis of a Single Polyprotein", pages 9-20, see entire document.
- Trends in Genetics, Volume 5, No. 2, issued February 1989, BAULCOMBE: "Strategies for Virus Resistance in Plants", pages 56-60, see entire document.

Description

FIELD OF THE INVENTION

[0001] This invention is directed to the production of plants with a reduced susceptibility to virus infection.

BACKGROUND OF THE INVENTION

[0002] Plant viruses are responsible for major losses in worldwide crop production. Much effort is directed towards the development of new plant varieties which exhibit increased resistance to viral infection. Until recently such efforts were primarily based on the traditional plant breeding approach, however this approach is often limited by a lack of sources of resistance within the crop species. The advent of modern molecular biology techniques has facilitated the development of new methods of rendering plant varieties resistant to virus attack that are not limited by a requirement for preexisting resistance genes within a species.

Molecular Approaches

[0003] Many of these molecular approaches are based on the theory of pathogen derived resistance (Sanford and Johnston, 1985). This theory predicts that a "normal" host (plant) - pathogen (virus) relationship can be disrupted if the host organism expresses essential pathogen derived genes. It has been proposed that host organisms expressing pathogen gene products in excess amounts, at an inappropriate developmental stage, or in a dysfunctional form may disrupt the normal replicative cycle of the pathogen and result in an attenuated or aborted infection of the host.

[0004] Two approaches typify this pathogen derived resistance: coat protein mediated resistance and antisense RNA expression. It has been demonstrated that transgenic plants expressing a plant virus coat protein can be resistant to infection by the homologous virus. This coat protein mediated resistance has been demonstrated for several virus groups. While the mechanism of this resistance is not yet fully understood, it has been suggested that the presence of the plant synthesized coat protein prevents the removal of the protein coat (uncoating) of an invading virus and/or virus movement within the infected plant, leading to resistance.

[0005] Plants which express an RNA molecule which is complementary to plus sense RNA species encoded by the virus may show a decreased susceptibility to infection by that virus. Such a complementary RNA molecule is termed antisense RNA. It is thought that the plant encoded antisense RNA binds to the viral RNA and thus inhibits its function.

Potyviruses

[0006] The Potato Virus Y, or potyvirus, family represents a large number of plant viral pathogens which collectively can infect most crop species including both monocotyledonous and dicotyledonous plants. Potyvirus infection can induce a variety of symptoms including leaf mottling, seed and fruit distortion and can severely compromise crop yield and/or quality (Hollings and Brunt, 1981).

[0007] Potyviruses have a single-strand plus sense RNA of circa 10,000 nucleotides which has a viral encoded protein linked to the 5' end and a 3' polyadenylate region. A single open reading frame codes for a 351 kDa polyprotein which is proteolytically processed into mature viral gene products. The RNA is encapsidated by approximately 2,000 copies of a coat protein monomer to form a virion. This capsid protein is encoded by the sequence present at the 3' end of the large open reading frame.

[0008] Potyviruses can be transmitted by aphids and other sap feeding insects and in some instances can also be transmitted in the seeds of infected plants. Replication of the viral RNA is thought to occur in the cytoplasm of infected plant cells after uncoating. The replication mechanism involves both translation of the plus sense RNA to yield viral gene products (which include a replicase and a proteinase) and also the synthesis of a minus sense RNA strand. This minus sense strand then acts as a template for the synthesis of many plus sense genomes which are subsequently encapsidated in coat protein to yield infectious mature "virions", thus complete the replicative cycle of the virus.

[0009] Experiments have been reported in which transgenic plants expressing the coat protein gene of a potyvirus show a reduced susceptibility to virus infection (Lawson *et al.* 1990; Ling *et al.* 1991; Stark and Beachy 1989).

[0010] EP-A-0242016 discloses the incorporation of genetic material, in particular cDNA corresonding to plant viral satellite RNA, into a plant such that, when the plant is infected by a plant virus, the expression of the incorporated material modifies the plant virus or its effects.

[0011] WO-A-9213090 discloses a method for producing transgenic plants with reduced virus susceptibility.

SUMMARY OF THE INVENTION

[0012] The disclosed invention concerns a method of producing plants with a decreased susceptibility to virus infection. This is achieved by transforming plants with a DNA molecule which includes a gene derived in part from the genome of a plant virus. This gene is specifically constructed to produce an untranslatable version of a plus sense RNA molecule required for viral replication. Thus expression of the gene within the plant causes the production of this non-functional molecule which then inhibits viral replication within the plant, rendering the plant resistant to viral infection.

[0013] In particular, invention provides an alternative and novel approach to rendering plants resistant to potyvirus infection.

[0014] Plants are transformed with a gene construct engineered to express an untranslatable form of the plus sense RNA which encodes the coat protein of a potyvirus.

[0015] In the case of Tobacco Etch Virus (TEV), it is demonstrated that tobacco plants transformed with such a gene construct accumulate the untranslatable plus sense RNA but do not produce detectable levels of the coat protein. It is further shown that these plants are resistant to TEV infection. It is also shown that tobacco cells expressing this untranslatable plus sense RNA do not support TEV replication, unlike control tobacco cells and also unlike tobacco cells which are engineered to express the plus sense translatable RNA and which, as a result, accumulate TEV coat protein. Although the exact mechanism is unknown, it is proposed that the untranslatable plus sense RNA inhibits viral replication by binding to the minus sense RNA and preventing the minus sense RNA from functioning in the replication cycle.

[0016] It is believed that this approach will be applicable to other potyviruses, to genes other than the coat protein gene and to other plus sense RNA virus families. It is also believed that this means of inhibiting gene function is applicable to other biological systems, including mammalian viruses.

DESCRIPTION OF DRAWINGS

[0017]

Fig. 1 represents the nucleotide sequence of the Tobacco Etch Virus genome and its deduced amino acid sequence, according to Allison et al. (1986). The nucleotide sequence of the plus sense strand of the DNA inserts is given. The first nucleotide (N) could not be determined unequivocally. The predicted amino acid sequence of the large ORF of reading frame three of the viron sense RNA is presented in the nucleotide sequence. This sequence is also set forth in SEQ ID No. 1 of the enclosed sequence listing. The termination codon at the end of the large ORF is marked with a *. The putative cleavage site between the large (54,000 *Mw)* nuclear inclusion protein and the capsid protein is indicated by the arrow. Oligonucleotide primer binding sites are underlined and labeled.

Fig. 2 is a schematic representation of the construction of pTC:FL, utilized in construction of transformation vectors for the invention. Restriction endonuclease sites were introduced into pTL 37/8595 at positions A, B and C in the diagram. Following these nucleotide changes the mutated pTL 37/8595 was digested with the restriction enzyme *NcoI,* the DNA fragment delineated by the restriction enzyme sites at B and C was removed, and the plasmid religated to generate pTC:FL. pTC:FL contains the Tobacco Etch Virus (TEV) coat protein nucleotide sequence flanked by *Bam*HI restriction sites and the TEV 5' and 3' untranslated sequences (UTS). T7 and SP6 promoters are also shown. Abbreviations used in this diagram are as follows: T7, T7 RNA polymerase promoter sequence; SP6, SP6 RNA polymerase promoter sequence; ori, origin of replication; M13 ori, bacteriophage M13 single-stranded origin of replication; amp^r, β-lactamase gene. Lightly stippled areas are TEV 5' and 3' untranslated sequences; solid black area, TEV genome cDNA nucleotides 144 to 200; striped area, a portion of the TEV NIb gene (TEV nt 8462-8517); heavily stippled areas, cDNA of TEV CP nucleotide sequence (TEV nt 8518-9309).

Fig. 3 is a schematic representation of the forms of the Tobacco Etch Virus coat protein gene inserted into tobacco in the invention. All constructs contained the enhanced CaMV 35S (Enh 35S) promoter, CaMV 35S 5' untranslated sequence (UTS) of 50 bp and the CaMV 35S 3' UTS/polyadenylation site of 110 bp. The nomenclature used to describe the transgenic plant lines is presented along with the gene products produced in those plant lines (far right column). Abbreviations are as follows: 35S, transgenic plants containing the CaMV 35S promoter and 5' and 3' UTS only; FL, transgenic plants containing the transgene coding for full-length, AS and RC transgenic plants contain the transgene expressed as an antisense form of the TEV CP gene, or an untranslated sense form of the TEV CP gene, respectively. Stippled areas represent various forms of the TEV CP nucleotide sequence.

Fig. 4 is a graphic representation of the appearance of systemic symptoms in plants infected with Tobacco Etch Virus showing responses of control plants and transformed plants generated as described in the invention. Ten B49 (wild type) plants and ten R2 plants of transgenic plant lines 35S #4, FL #3, FL #24, homozygous for the inserted TEV gene, were mechanically inoculated with 50 µl of 1:10 dilution of infected plant sap (A). Twenty B49 plants and 20 R1 plants of lines AS #3 and RC #5 were mechanically inoculated with 50 µl of 5 µg/ml TEV (B).

Plants were examined daily for the appearance of systemic symptoms. Plants were evaluated daily, and any plant displaying systemic symptoms (attenuated or wild-type) were recorded as symptomatic.

SEQUENCE LISTING

[0018] The attached sequence listing sets forth nucleotide sequences relevant to the present invention.

[0019] SEQ ID No. 1 is the complementary DNA sequence corresponding to the Tobacco Etch Virus Genome.

[0020] SEQ ID No. 2 is the nucleotide sequence of the modified Tobacco Etch Virus coat protein gene present in pTC:FL.

[0021] SEQ ID No. 3 is the nucleotide sequence of the modified Tobacco Etch Virus coat protein gene present in pTC:RC.

[0022] SEQ ID No. 4 is the nucleotide sequence of the modified Tobacco Etch Virus coat protein gene present in pTC:AS. It is the inverse complement of SEQ ID No. 2.

DETAILED DESCRIPTION

[0023] The present invention relates to genetically engineered plants which are transformed with a DNA molecule encoding an untranslatable plus sense RNA molecule.

Definition of Terms

[0024] Susceptible plant: A plant that supports viral replication and displays virus-induced symptoms.

[0025] Resistant plant: A plant wherein virus-induced symptoms are attenuated and virus replication is attenuated.

[0026] Plus sense RNA (and sense RNA): That form of an RNA which can serve as messenger RNA.

[0027] Minus sense RNA: That form of RNA used as a template for plus sense RNA production.

[0028] Antisense RNA: RNA complementary to plus sense RNA form.

[0029] R_o generation: Primary transformants.

[0030] R_1 generation: Progeny of primary transformants.

[0031] R_2 generation: Second generation progeny of R_o generation (i.e., progeny of R_1 generation).

[0032] A gene derived in part from a plant virus RNA molecule: At least the portion of the gene encoding the untranslatable RNA molecule is derived from a plant virus RNA molecule.

GENERAL DESCRIPTION

[0033] An untranslatable plus sense RNA molecule is encoded by a gene located on the DNA molecule. The gene comprises DNA derived from a plant virus RNA genome and also DNA from heterologous sources. The DNA from heterologous sources includes elements controlling the expression of the virus-derived DNA sequences. The DNA sequence of the gene is specifically altered so as to render the RNA molecule transcribed from the gene untranslatable. The presence of this untranslatable plus sense RNA within the cells of the transformed plant reduces the susceptibility of the plant to viral infection.

[0034] More particularly, the portion of the gene which comprises DNA from a plant virus has been derived from a potyvirus. Plants transformed with the DNA molecule containing the gene are less susceptible to infection by potyviruses. Most specifically, the DNA from the potyvirus source has been derived from the coat protein gene of Tobacco Etch Virus and transformed plants are resistant to infection by Tobacco Etch Virus. Plants which can be made resistant to potyvirus infection include, but are not limited to, tobacco.

[0035] Accordingly, the present invention provides a method for genetically engineering plants by insertion, into the plant genome, a DNA construct containing a recombinant gene derived from a potyvirus genome such that the engineered plants display resistance to the potyvirus.

[0036] In accordance with one aspect of the presen invention, genetically transformed plants which are resistant to infection by a plant potyvirus are produced by inserting into the genome of the plant a DNA sequence which causes the production of an untranslatable coat protein RNA of the potyvirus.

[0037] In accordance with another aspect of the present invention, a DNA sequence is provided to function in plant cells to cause the production of an untranslatable plus sense RNA molecule. There has also been provided, in accordance with yet another aspect of the present invention, bacterial and transformed plant cells that contain the above-described DNA. In accordance with yet another aspect of the present invention, a differentiated tobacco plant has been provided that comprises transformed tobacco cells which express the untranslatable coat protein RNA of Tobacco Etch Virus and which plants exhibit resistance to infection by Tobacco Etch Virus.

[0038] A mechanism by which an untranslatable plus sense RNA molecule, such as described in the current invention

can function to inhibit the normal biological function of a minus sense RNA molecule is proposed. One skilled in the art will recognize that the novel approach described herein is not limited to the specific experimental example given and will appreciate the wider potential utility of the invention.

[0039] The expression of a plant gene which exists in double-stranded DNA form involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3' nontranslated region which causes polyadenylate nucleotides to be added to the 3' end of the viral RNA. Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the "promoter." The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding strand of RNA.

[0040] A number of promoters which are active in plant cells have been described in the literature. Promoters which are known or are found to cause transcription of viral RNA in plant cells can be used in the present invention. Such promoters may be obtained from plants or viruses and include, but are not limited to, the CaMV 35S promoter. As described below, it is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of untranslatable plus sense RNA to render the plant substantially resistant to virus infection. The amount of untranslatable plus sense RNA needed to induce resistance may vary with the plant type. Accordingly, while the 35S promoter is preferred, it should be understood that this promoter may not be the optimal one for all embodiments of the present invention. Furthermore, the promoters used in the DNA constructs of the invention may be modified, if desired, to affect their control characteristics. DNA sequences have been identified which confer regulatory specificity on promoter regions. For example, the small subunit of the ribulose bis-phosphate carboxylase (ss RUBISCO) gene is expressed in plant leaves but not in root tissues. A sequence motif that represses the expression of the ss RUBISCO gene in the absence of light, to create a promoter which is active in leaves but not in root tissue, has been identified. This and/or other regulatory sequence motifs may be ligated to promoters such as the CaMV 35S promoter to modify the expression patterns of a gene. Chimeric promoters so constructed may be used as described herein. For purposes of this description, the phrase "CaMV 35S promoter" will therefore include all promoters derived by means of ligation with operator regions, random or controlled mutagenesis, as well as tandem or multiple copies of enhancer elements, and the like.

[0041] The 3' nontranslated region of genes which are known or are found to function as polyadenylation sites for viral RNA in plant cells can be used in the present invention. Such 3' nontranslated regions include, but are not limited to, the 3' transcribed, nontranslated region of the CaMV 35S gene and the 3' transcribed, nontranslated regions containing the polyadenylation signals of the tumor-inducing (TI) genes of *Agrobacterium*, such as the tumor morphology large (tml) gene. For purposes of this description, the phrase "CaMV 35S 3' nontranslated region" will therefore include all such appropriate 3' nontranslated regions.

[0042] The DNA constructs of the disclosed embodiment contain, in double-stranded DNA form, a portion of a cDNA version of the single-stranded RNA genome of TEV. In potyviruses, including TEV, the viral genome includes genes encoding the coat protein, a replicase enzyme and a proteinase. The disclosed embodiment utilizes the region of the genome encoding the coat protein gene. In considering the present invention and the evidence for the proposed mechanism by which an untranslatable plus sense RNA molecule can inhibit viral replication, those skilled in the art will recognize that other portions of a potyvirus genome could be substituted for the coat protein gene. Furthermore, it will be apparent that suitable genomic portions are not limited to complete gene sequences.

[0043] A disclosed embodiment of the invention utilizes a double-stranded complementary DNA (cDNA) derived from the region of the TEV genome encoding the coat protein gene. To the 5' end of this cDNA is ligated the CaMV 35S promoter and CaMV 35S RNA 5' nontranslated region. To the 3' end is ligated the CaMV 35S 3' nontranslated region. These 5' and 3' sequences are present to cause transcription of the gene in plant cells by the cellular enzyme RNA polymerase to produce an RNA molecule of sequence corresponding to the sequence of the coat protein cDNA sequence. Ordinarily, such an RNA would then be translated by ribosomes which would synthesize a protein of amino acid sequence specified by the nucleotide sequence of the RNA molecule. Particular amino acids are specified by nucleotide triplets termed codons. Codons which stipulate translation initiation and termination are also present in DNA and RNA sequences. The current invention relates to RNA molecules which are untranslatable by ribosomes. In the preferred embodiment the sequence of the TEV cDNA encoding the coat protein is mutated by a standard *in vitro* mutagenesis technique to produce a frameshift mutation early in the coat protein structural gene immediately followed by three translation termination signal codons. These mutations do not affect the ability of RNA polymerase to transcribe an RNA molecule from the cDNA but prevent translation of the transcribed RNA by ribosomes. Those skilled in the art will recognize that for the disclosed gene and for other genes, DNA sequences can be altered in other ways to cause the DNA to encode an untranslatable plus sense RNA molecule. Thus the disclosed invention is not limited to the mutations disclosed.

[0044] A disclosed embodiment utilizes a cDNA encoding the coat protein gene of TEV, mutated so as to encode an untranslatable plus sense RNA. It will be obvious to one skilled in the art that further sequence alteration of the cDNA

molecule could be used to confer additional features on the untranslatable plus sense RNA molecule. Additional features include those which would result in increased viral resistance of plants transformed with the cDNA molecule encoding an untranslatable plus sense RNA. The inclusion of a ribozyme sequence which causes the RNA catalyzed destruction of the target RNA molecule would constitute one such additional feature. Suitable ribozyme sequences are known, as discussed in Tabler and Tsagris (1991).

[0045] A DNA construct in accordance with the present invention is introduced, via a suitable vector and transformation method as described below, into plant cells and plants transformed with the introduced DNA are regenerated. Various methods exist for transforming plant cells and thereby generating transgenic plants. Methods which are known or are found to be suitable for creating stably transformed plants can be used in this invention. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome mediated transformation; polyethylene mediated transformation; transformation using viruses; microinjection of plant cells; microprojectile bombardment of plant cells and *Agrobacterium tumefaciens* (AT) mediated transformation. The latter technique is the method of choice for the disclosed preferred embodiment of the present invention.

[0046] In an embodiment of the current invention, the DNA sequences comprising the CaMV 35S promoter and CaMV 35S nontranslated 3' region and the mutated cDNA encoding an untranslatable plus sense RNA derived from the TEV coat protein gene are combined in a single cloning vector. This vector is subsequently transformed into AT cells and the resultant cells are used to transform cultured tobacco cells.

[0047] Vectors suitable for the AT mediated transformation of plants with the DNA of the invention are disclosed. It will be obvious to one skilled in the art that a range of suitable vectors is available, including those disclosed by Bevan (1983), Herrera-Estrella (1983), Klee (1985) and EP-A-120516 (Schilperoort et al.). Suitable vectors are available on a commercial basis from Clontech (Palo Alto, CA) and Pharmacia LKB (Pleasant Hill, CA) and other sources.

[0048] Following the transformation of plant cells and regeneration of transformed plants with the DNA molecules as described, regenerated plants are tested for increased virus resistance. Plants are preferably exposed to the virus at a concentration within a range where the rate of disease development correlates linearly with virus concentration. Methods for virus inoculation are well known to those skilled in the art and are reviewed by Kado and Agrawai (1972). One such method includes abrading a leaf surface with an aqueous suspension containing an abrasive material such as carborundrum and virus or dusting leaves with such an abrasive material and subsequently applying the virus onto the leaf surface. A virus suspension can be directly inoculated into leaf veins or alternatively plants can be inoculated using insect vectors. The virus suspension may comprise purified virus particles, or alternatively, sap from virus infected plants may be utilized.

[0049] Transformed plants are then assessed for resistance to the virus. The assessment of resistance or reduced susceptibility may be manifest in different ways dependant on the particular virus type and plant type. Those skilled in the art will realize that a comparison of symptom development on a number of inoculated untransformed plants with symptom development on similarly inoculated transformed plants will provide a preferred method of determining the effects of transformation with the specified DNA molecule on plant resistance. Symptoms of infection include, but are not limited to leaf mottling, chlorosis and etching. Plants showing increased viral resistance may be recognized by delay in appearance of such symptoms or attenuation or total lack of such symptoms.

Example

[0050] Work with tobacco plants and the Tobacco Etch Virus (TEV) is illustrative of the invention.

Construction of gene encoding untranslatable plus sense RNA molecule.

[0051] The Highly Aphid Transmissible (HAT) isolate of Tobacco Etch Virus (TEV) was obtained from Dr. Tom Pirone (University of Kentucky) and maintained in *Nicotiana tabacum* (Burley 21). The virus was purified from *Nicotiana tabacum* (Burley 21) 20 to 30 days following inoculation. Viral purification and RNA isolation procedures have been described (Dougherty and Hiebert (1980a). Complementary DNA (cDNA) was synthesized, made double-stranded and inserted into the bacterial plasmid pBR322 as described by Allison et al. (1985a, 1985b, 1986). cDNA synthesis was accomplished as follows: Purified viral RNA primed with oligo(dT_{12-18}) served as a template for single-strand cDNA synthesis by reverse transcriptase. Following the addition of homopolymeric tracts of deoxycytidine 5' monophosphate, second-strand synthesis, primed with oligo(dG_{12-18}), was completed with DNA polymerase I. *Sal*I and *Eco*RI linkers were ligated to the double-stranded cDNA and inserted into the bacterial plasmid pBR322 (Kurtz and Nicodemus 1981). The resulting cDNA clones were screened by colony hybridization (Hanahan and Meselson 1980) with oligo(dT_{12-18}) primed, ^{32}P-labeled single-stranded TEV cDNA. Plasmid DNA was isolated from colonies which hybridized with the probe, and the *Sal*I/*Eco*RI cDNA inserts were sized by electrophoresis in a 0.8% (w/v) agarose gel using a horizontal water-cooled gel apparatus.

[0052] The *Sal*I/*Eco*RI inserts from the recombinant molecules were isolated from an agarose gel with NA45 membrane (Schleicher & Schuell, Keene, NH) according to the manufacturer's protocol. The following restriction enzymes were used either alone or in combination to digest the isolated cDNA insert: *Hind*III, *Xho*I, *Alu*I, *Hae*III, *Rsa*I, *Sau*3A, and *Taq*I. Restriction enzyme digestion products were inserted into the DNA of an appropriate M13 bacteriophage (Messing 1983) seiecred for the presence of corresponding polylinker restriction sites, and their nucleotide sequences were determined by dideoxy chain termination.

[0053] Plasmid pTL 37/8595 (Carrington and Dougherty 1987; Carrington et al. 1987) contains a cDNA copy of the genomic sequence of HAT TEV corresponding to nucleotides (nt) 1-200 and nt 8462-9495 (Fig. 2). (Numbering of the TEV genome nucleotides is according to that presented in Allison et al. 1986). The nucleotide sequence and deduced amino acid sequence of the Tobacco Etch Virus genome and the numbering system utilized by Allison et al. (1986) and herein is shown in Fig. 1 and SEQ ID No. 1 in the attached sequence listing. The first and last codons of the coat protein (CP) coding region in the TEV genome are nt 8518-8520 (encoding the amino acid serine) and 9307-9309 (opal stop codon) respectively. pTL 37/8595 was subject to *in vitro* site-directed mutagenesis as described by Taylor et al. (1985a, 1985b). In all cases, nucleotide changes were confirmed by dideoxy-nucleotide sequencing (Sanger et al. 1977).

[0054] TEV nt 9312-9317 were first mutated (Fig. 2) to generate a *Bam*HI restriction site (GGATCC). TEV nt 8516-8521 were then altered to generate an *Nco*I site (CCATGG), changing the first codon of the TEV CP coding region from AGT (Ser), to ATG (Met). A single oligonucleotide was then used to mutate TEV nt 133-138 to a *Bam*HI restriction site (GGATCC), nt 143-148 to an NcoI restriction site (CCATGG) and nt 142 to a deoxyadenylate residue. These mutations generated an NcoI site centered on the first codon of the TEV ORF and in a good translational start context as described by Kozak (1984). Digestion of the resulting plasmid with the restriction enzyme *Nco*I; removing TEV nt # 143-200/8462-8516, and religation generated plasmid pTC:FL. pTC:FL contained only the TEV CP gene flanked by *Bam*HI restriction sites and TEV 5' and 3' untranslated sequences (see Fig. 2). The nucleotide sequence of the TEV CP gene in pTC:FL produced by this mutagenesis scheme is shown in SEQ ID No. 2 in the attached sequence listing.

[0055] Plasmid pTC:RC (RNA Control, producing untranslatable plus sense RNA) was generated by insertion of a single deoxythymidylate residue after TEV nt 8529, and point mutations of TEV nt 8522 (G to C), 8534 (C to A), 8542 (G to A), and 8543 (A to G) to create a frameshift mutation immediately followed by three stop codons. An *Nhe*I restriction site (GCTAGC) was simultaneously generated, for screening purposes, at nt 8539-8544. The nucleotide sequence of the TEV CP gene in pTC:RC produced by this mutagenesis scheme is shown in SEQ ID No. 3 in the attached sequence listing.

[0056] All plasmids described above were linearized with *Hind*III, transcribed with T7 RNA polymerase (Melton et al. 1984), and translated in a rabbit reticulocyte lysate containing ^{35}S Methionine (Dougherty and Hiebert 1980a). Radiolabeled translation products were analyzed by electrophoretic separation on a 12.5% acrylamide gel containing SDS (Laemmli 1970) and detected by autoradiography. Transcripts of plasmid pTC:RC produced no detectable protein products, while transcripts from pTC:FL produced proteins of the expected sizes.

[0057] The various forms of the CP nucleotide sequence were then inserted as *Bam*HI cassettes into the plant expression vector pPEV (see below and Fig. 3).

[0058] The full length TEV CP open reading frame of pTC:FL was inserted in the reverse orientation to make the antisense (AS) construct pTC:AS. The nucleotide sequence of the TEV CP gene in pTC:AS is shown in SEQ ID No. 4 in the attached sequence listing.

Transformation Vector Construction

[0059] Construction of pPEV. The vector pPEV is part of a binary vector system for *Agrobacterium tumefaciens* mediated plant cell transformation. Plasmid pPEV was constructed from the plasmids pCGN 2113 (Calgene), pCIB 710 and pCIB 200 (Ciba Geigy Corp.). pCGN 2113 contains the "enhanced" Cauliflower Mosaic Virus (CaMV) 35S promoter (CaMV sequences -941 to 90/-363 to +2, relative to the transcription start site) in a pUC derived plasmid backbone. pCIB 710 has been described (Rothstein et al. 1987) and pCIB 200 is a derivative of the wide host range plasmid pTJS 75 (Schmidhauser and Helinski 1985) which contains left and right *A. tumefaciens* T37 DNA borders, the plant selectable NOS/NPT II chimeric gene from the plasmid Bin 6 (Bevan 1984) and part of a pUC polylinker. The small *Eco*RI-*Eco*RV DNA fragment of pCIB 710 (Rothstein et al. 1987) was ligated into *Eco*RI-*Eco*RV digested pCGN 2113. This regenerated the enhanced CaMV 35S promoter (Kay et al. 1987) of pCGN 2113 and introduced the CaMV 35S 5' and 3' untranslated sequences into pCGN 2113. The CaMV 35S promoterterminator cassette of the resulting plasmid was isolated as an *Eco*RI-*Xba*I DNA fragment and ligated into *Eco*RI-*Xba*I digested pCIB 200 to generate pPEV. CP nucleotide sequences from PTC:FL, pTC:RC, and pTC:AS were cloned as *Bam*HI cassettes into *Bam*HI digested pPEV and orientation of inserts confirmed by digestion with appropriate restriction endonucleases.

Transformation and Regeneration of Tobacco

[0060] pPEV plasmids containing TEV CP ORFs were mobilized from *E. coli* HB101 into *A. tumefaciens* A136 containing plasmid pCIB 542 (Ciba Geigy), using the helper plasmid pRK 2013 in *E. coli* HB101 and the tri-parental mating system of Ditta et al. (1980). Plasmid pCIB 42 supplied *vir* functions necessary for T-DNA transfer.

[0061] Leaf discs of *Nicotiana tabacum cv* Burley 49 were transformed and whole plants regenerated according to Horsch et al. (1985). Transformed tissue was selected by culturing callus on MS plates (Murashige and Skoog 1962) containing 1 μ g/ml 6-benzylaminopurine (Sigma Corp.), 01 μg/ml α-naphthaleneacetic acid (Sigma Corp.), 500 μg/ml carbenicillin and 100 μg/ml Kanamycin sulfate (Sigma Corp.). Shoots were rooted on MS plates containing 500 μg/ml carbenicillin and 100 μg/ml kanamycin sulfate, and plantlets were transplanted into soil and transferred directly into the greenhouse approximately 2-3 weeks after rooting.

[0062] R0, R1 and R2 generation plants were screened by western and/or northern blot analyses. R2 seed (ca. 100 seeds per R2 plant) was screened for the kanamycin-resistant phenotype (kan^r) by surface sterilizing seed in 10% bleach for 5 min., washing twice in sterile water and germinating on MS plates containing 100 μg/ml kanamycin sulfate. R2 seed lines which were 100% kanamycin resistant were screened by western blot analysis for expression of TEV coat protein. Those transgenic plant lines generated and their nomenclature are presented in Fig. 3.

Molecular Analyses of Transgenic Plants

[0063] Transgenic tobacco plants were analyzed by western and northern blot analyses to determine the nature of protein and RNA products produced respectively. Total RNA samples isolated from the various transgenic lines were analyzed in northern blot hybridization studies. Total nucleic acids were isolated from tissue and RNA precipitated with LiCl as described by Verwoerd et al. (1989). RNAs were electrophoretically separated on 1.2% agarose gels containing 6% (v/v) formaldehyde and transferred to nitrocellulose. Prehybridization and hybridization conditions were as described in Sambrook et al. (1989). Strand specific riboprobes were generated from SP6 or T7 DNA dependent RNA polymerase transcription reactions of pTL 37/8595 linearized with the restriction enzymes Asp718 (Boehringer Mannheim, Indianapolis, IN) or *Hind*III, respectively, using α-labelled ^{32}P-CTP ribonucleotide and suggested procedures (Promega, Madison, WI).

[0064] An RNA transcript of approximately 1,000 nt was expected with all transgenic plant lines. Such a TEV CP transcript was detected in CP expressing plant lines by using a minus sense riboprobe containing the TEV CP sequence. A similar transcript was detected in AS plants by using a plus sense riboprobe containing the TEV CP sequence. The transcript in the RC line, while detected with a minus sense riboprobe, may have migrated as a slightly larger (ca 1,100-1,200 nt) RNA species, possibly due to termination at an alternately selected site and/or a longer poly-A tail on the transcript. Differing levels of CP transcript accumulation were observed among different transgenic plant lines. Transgenic plant lines expressing the coat protein of TEV were identified by western blot analysis using polyclonal antisera to TEV CP. Tissue samples of regenerated plants were ground in 10 volumes of 2X Laemmli (Tris-glicine) runner buffer (Laemmli 1970) and clarified by centrifugation in a microcentrifuge for 10 min. at 10,000xg. Protein concentration was estimated by the dye binding procedure of Bradford (1976) using BSA as a standard. Protein samples (50 μg total protein) were separated on a 12.5% polyacrylamide gel containing SDS and subjected to the immunoblot transfer procedures described by Towbin et al. (1979). Anti-TEV coat protein polyclonal primary antibodies, alkaline phosphatase conjugated secondary antibodies and the chromogenic substrates NBT (para-nitro blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3-indoyl phosphate para-toluidine salt) were used to detect bound antigen.

[0065] Coat protein products produced in FL plants were stable and accumulated to different levels in individual transgenic plant lines. It was estimated by western blot analysis that between 0.01% to 0.001% of total extracted protein was TEV CP.

Assessment of Resistance to TEV

[0066] Eight-week-old (circa 15 cm tall) R1 and R2 plants were inoculated with either purified virus preparations or infected plant sap. Inoculum was applied with sterile, premoistened cotton swabs. Infected plant sap inoculum was prepared by grinding TEV-infected *N. tabacum* Burley 21 leaf tissue (2 weeks postinoculation) in carborundum and 50 mM sodium phosphate buffer (pH 7.8) at a ratio of 1gm:02gm:10mls, respectively, and filtering the homogenate through cheesecloth. TEV virons were purified as described by Dougherty and Hiebert (1980b). One leaf per plant was dusted lightly with carborundum (320 grit) and inoculated at two interveinal locations with 50 μl (total) of inoculum. Inoculated plants were examined daily and the appearance and severity of systemic symptoms recorded. Symptoms on any leaf above the inoculated leaf were considered to be systemic.

[0067] Typically, inoculation of Burley 49 plants with TEV (either purified virus or plant sap) resulted in severe chlorosis and mosaic and mottle on systemically infected leaves approximately 6-7 days after inoculation. Severe etching of the

leaf followed within a few days. It was observed that transgenic plants containing only the CaMV promoter and untranslated sequences (i.e., 35S plant line) responded to challenge inoculation in a manner similar to wild type Burley 49, developing extensive chlorosis and etching at the same rate (Fig. 4A). Plant lines which expressed FL TEV CP showed little or no delay in the appearance of symptoms when inoculated with infected plant sap. However, FL transgenic plants did show a slight attenuation of symptoms and eventually (2-4 weeks after initial appearance of symptoms), younger leaf tissue emerged devoid of symptoms and virus as demonstrated by back inoculation experiments. Typically chlorosis and etching on older systemic leaves was limited.

[0068] Ten independently transformed RC lines and seven independently transformed AS lines were obtained. Progeny from three of the RC lines, including line RC #5 and from one of the AS lines, including AS #3, showed an altered response to viral infection relative to control plants. All of these lines were verified to be transformed and were producing expected RNA products. A possible explanation for the variation in observed phenotype is the previously noted "position effect" whereby the expression of genes from identical DNA sequences integrated at different locations within the genome show varying patterns of tissue specificity.

[0069] Ten R2 expressing plants of the FL expressing line were inoculated with infected plant sap, and 20 R1 plants of lines AS #3 and RC #5 were inoculated with 50 µl of a 5 µg/ml solution of purified TEV. Identical results to those obtained by purified TEV inoculation were obtained when AS #3 and RC #5 R1 plants were inoculated with TEV-infected plant sap, as described above.

[0070] Transgenic Burley 49 plant lines AS #3 and RC #5, expressing only TEV CP related RNA sequences, showed a delay in the appearance of symptoms and a modification of symptoms when inoculated with TEV (Fig. 4B). Since the 20 R1 plants were not screened for expression of CP RNA prior to inoculation, some of the symptomatic plants represented non-expressing plants in which the gene of interest had been lost during Mendelian segregation. Modified symptoms on AS #3 plants appeared as small chlorotic lesions often associated with a vein. Most of the leaves were devoid of symptoms and virus (determined by back inoculation experiments). Approximately 15% of RC #5 plants showed symptoms which were identical to those of infected Burley 49. However, the remaining RC #5 plants were entirely asymptomatic, and virus was not detected in back inoculation studies.

[0071] Plants from TEV resistant AS and RC lines showed no increased resistance, relative to untransformed controls, to infection by two other members of the potyvirus family, namely Tobacco Vein Mottling Virus and Potato Virus Y.

[0072] R_2 generation plants derived from TEV-resistant RC plants showed the expected Mendelian pattern of inheritance of the TEV-resistant phenotype.

Analysis of TEV Replication in Protoplasts Derived from Transgenic Plant Lines

[0073] In an attempt to explain the results obtained when AS and RC transgenic plants were challenged with TEV, it was sought to determine if all of the transgenic plant lines would support virus replication at a level comparable to Burley 49. Accumulation of viral encoded proteins was used as an indirect indicator of viral replication. Protoplasts were derived from leaf tissue of homozygous CP expressing plants and electroporated according to the procedure of Luciano et al. (1987) with TEV RNA. Protoplasts were prepared from transgenic plants and electroporated according to the procedure of Luciano et al. (1987). Protoplasts (1×10^6) were resuspended in 450 µl electroporation buffer (330 mM mannitol, 1 mM KPO_4 pH 7.0, 150 mM KCl) and electroporated using a BTX Transfector 300 (BTX San Diego, CA) (950 micro Farads, 130-volt pulse amplitude, 3.5 mm electrode gap) in the presence or absence of 6 µg of purified TEV RNA. After electroporation, protoplasts were incubated for 96 hours in incubation medium as described in Luciano et al. (1987). Protoplasts were extracted in 2X Laemmli (Trisglycine) running buffer, and 5×10^4 extracted protoplasts were then subjected to western blot analysis as described above. Protoplast viability was measured by dye exclusion as described in Luciano et al. (1987). All electroporated protoplast samples had equivalent viability counts. The results indicated that protoplasts from all FL plant lines supported virus replication at levels comparable to wild type Burley 49 protoplasts. R1 transgenic plants from lines AS #3 and RC #5 were initially screened by northern analysis, and leaves from positive expressors were used in the production of protoplasts. Transfected protoplasts derived from AS #3 plants supported TEV replication, albeit at a reduced level. Protoplasts derived from RC #5 transgenic plant leaf tissue did not support TEV replication at a detectable level. These results, and those presented in the whole plant inoculation series, suggested AS and RC plants interfere with TEV replication.

Discussion of Data

[0074] The above example indicates that varying degrees of protection from TEV infection can be achieved by overexpression of coat protein and by expression of an antisense RNA. The current invention which comprises the expression of an untranslatable plus sense RNA molecule provides protection against TEV infection that is more effective than either of these two methods. Plants of line RC #5, transformed with the disclosed DNA molecule encoding an untranslatable plus sense RNA derived from the TEV coat protein gene, were asymptomatic and appear to be com-

pletely protected from virus infection. The disclosed invention therefore represents a new and effective way of generating potyvirus resistant germplasm.

[0075] Tobacco protoplasts derived from plants expressing the antisense RNA supported a reduced level of TEV replication compared to control cells derived from untransformed plants. In contrast, tobacco protoplasts derived from plants of line RC #5, expressing the untranslatable plus sense RNA did not support detectable TEV replication. This suggests that the untranslatable plus sense RNA was more effective at blocking TEV replication in the cells of those transformed plants tested.

[0076] It is proposed that the untranslatable plus sense RNA inhibits viral replication by hybridizing to the minus sense RNA replicative template of TEV. The finding that plants expressing untranslatable plus sense RNA derived from the TEV coat protein gene are not protected from infection by Potato Virus Y or Tobacco Vein Mottling Virus is therefore explained by the circa 40-50% amino acid sequence divergence between the coat proteins of these viruses and TEV (Allison et al. 1986; Robaglia et al. 1989; Domier et al. 1986).

[0077] From the above-described findings, it would be reasonable and entirely predictable that if plants were transformed with a gene encoding an untranslatable plus sense RNA derived from a gene which was highly conserved between viruses of the potyvirus family, that these plants would be protected from infection by a wide range of viruses. Regions of the potyvirus genome which are sufficiently conserved between potyvirus types to be potentially useful in such an approach may be readily determined by one skilled in the art. Highly conserved regions may be determined by reference to published sequence data (Allison et al. 1986; Robaglia et al. 1989; Domier et al. 1986; Lain et al. 1989; Maiss et al. 1989). The utility of the identified regions could be readily determined using the methodologies described above and substituting the defined region for the TEV coat protein gene.

[0078] Regions of the potyvirus genome potentially suitable include, but are not limited to the genes encoding the viral replicase and the viral proteinase. Furthermore, it will be apparent to one skilled in the art that highly conserved portions of a particular gene may also serve in this role.

[0079] It will also be apparent to one skilled in the art that the described invention may also be used to produce plants resistant to viruses outside of the potyvirus family in instances where these viruses also produce a minus sense RNA replicative template.

BIBLIOGRAPHY

[0080] Allison et al. 1985a. Virology 147:309-316. Allison et al. 1985b. Proc. Natl. Acad. Sci. USA 82:3969-3972.

[0081] Allison et al. 1986. Virology 154:9-20.

[0082] Bevan 1984. Nucl. Acids Res. 12:8711-8721.

[0083] Bradford 1976. Nucl. Acids Res. 12:8711-8721.

[0084] Carrington and Dougherty 1987. J. Virol. 61:2540-2548.

[0085] Carrington et al. 1987. Nucl. Acids Res. 15:10066.

[0086] Ditta et al. 1980. Proc. Natl. Acad. Sci. USA 77:7347-7351.

[0087] Domier et al. 1986. Nucl. Acids Res. 14:5417-5430.

[0088] Dougherty and Hiebert 1980a. Virology 101:466-474.

[0089] Dougherty and Hiebert 1980b. Virology 104:183-194.

[0090] Hanahan and Meselson 1980. Gene 10:63-67.

[0091] Herrera-Estrella et al. 1983. Nature (London) 303:209-213.

[0092] Hollings and Brunt 1981. Commonwealth Mycological Institute/Association of Applied Biologists Descrip. Plant Viruses, No. 245.

[0093] Horsch et al. 1985. Science 227:1229-1231.

[0094] Kado and Agrawai 1972. Principles and Techniques in Plant Virology.

[0095] Kay et al. 1987. Science 236:1299-1302.

[0096] Kozak 1984. Nature (London) 308:241-246.

[0097] Kurtz and Nicodemus 1981. Gene 13:145-152.

[0098] Laemmli 1970. Nature (London) 227:680-685.

[0099] Lain et al. 1989. Virus Res. 13:157-172.

[0100] Lawson et al. 1990. Bio/Technology 8:127-134.

[0101] Ling et al. 1991. Bio/Technology 9:752-758.

[0102] Luciano et al. 1987. Plant Science 51:295-303.

[0103] Maiss et al. 1989. J. Gen. Virol. 70:513-524.

[0104] Melton et al. 1984. Nucl. Acids Res. 12:7145-7156.

[0105] Messing 1983. "Methods in Enzymology" (R. Wu, L. Grossman, and K. Moldave, eds.). Vol. 101c, pp. 20-78. Academic Press, New York.

[0106] Murashige and Skoog 1962. Plant Physiol. 15:473-497.

[0107] Robaglia et al. 1989. J. Gen. Virol. 70:935-947.
[0108] Rothstein et al. 1987. Gene 53:153-161.
[0109] Sambrook et al. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratories, Cold Spring Harbor, New York.
[0110] Sanford and Johnston 1985. J. Theor. Biol. 113:395-405.
[0111] Sanger et al. 1977. Proc. Natl. Acad. Sci. USA 74:5463-5467.
[0112] Schmidhauser and Helinski 1985. J. Bact. 164:446-455.
[0113] Stark and Beachy 1989. Bio/Technology 7:1257-1262.
[0114] Tabler and Tsagris 1991. Gene 100:175-183.
[0115] Taylor et al. 1985a. Nucl. Acids Res. 13:8749-8764.
[0116] Taylor et al. 1985b. Nucl. Acids Res. 13:8765-8785.
[0117] Towbin et al. 1979. Proc. Natl. Acad. Sci. USA 76:4350-4354.
[0118] Verwoerd et al. 1989. Nucl. Acids Res. 17:2372.

SEQUENCE LISTING

[0119]

(1) GENERAL INFORMATION:

(i) APPLICANT: William G. Dougherty and John A. Lindbo

(ii) TITLE OF INVENTION: Production of Plants Showing Immunity to Viral Infection via Introduction of Genes Encoding Untranslatable Plus Sense RNA Molecules

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Richard J. Polley

(B) STREET: One World Trade Center 121 S.W. Salmon Street, Suite 1600

(C) CITY: Portland

(D) STATE: Oregon

(E) COUNTRY: United States of America

(F) ZIP: 97204

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 5.25 inch

(B) COMPUTER: IBM PC Compatible

(C) OPERATING SYSTEM: MS DOS

(D) SOFTWARE: WordPerfect 5.1

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 07/838,509

(B) FILING DATE: February 19, 1992

(C) CLASSIFICATION: 435

(vi) PRIOR APPLICATION DATA: None

(vii) ATTORNEY/AGENT INFORMATION

(A) NAME: Richard J. Polley, Esq.

(B) REGISTRATION NUMBER: 28,107

(C) REFERENCE/DOCKET NUMBER: 245-35829/RJP

(viii) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (503) 226-7391

(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9495

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: cDNA to genomic RNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(v) FRAGMENT TYPE: N/A

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Tobacco Etch Virus (TEV)

(B) STRAIN: Highly Aphid Transmitted (HAT)

(vii) IMMEDIATE SOURCE: TEV propagated in N. tabacum Burley 49

(viii) POSITION IN GENOME: N/A

(ix) FEATURE:

(A) NAME/KEY: Coat protein gene

(B) LOCATION: Genomic nucleotides 8518-9306

(C) IDENTIFICATION METHOD: --

(D) OTHER INFORMATION: SEQ. ID No. 1 is the cDNA corresponding to the Tobacco Etch Virus Genome.

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Allison et al.

(B) TITLE: The nucleotide sequence of the coding region of Tobacco Etch Virus Genomic RNA: Evidence for the Synthesis of a Single Polyprotein

(C) JOURNAL: Virology

(D) VOLUME: 154

(E) ISSUE: --

(F) PAGES: 9-20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
NAAATAACAA ATCTCAACAC AACATATACA AAACAAACGA ATCTCAAGCA ATCAAGCATT      60

CTACTTCTAT TGCAGCAATT TAAATCATTT CTTTTAAAGC AAAAGCAATT TTCTGAAAAT     120

TTTCACCATT TACGAACGAT AGCA ATG GCA CTG ATC TTT GGC ACA GTC AAC GCT     174
                           Met Ala Leu Ile Phe Gly Thr Val Asn Ala
                             1               5                  10

AAC ATC CTG AAG GAA GTG TTC GGT GGA GCT CGT ATG GCT TGC GTT ACC       222
Asn Ile Leu Lys Glu Val Phe Gly Gly Ala Arg Met Ala Cys Val Thr
                 15                  20                  25
```

```
AGC GCA CAT ATG GCT GGA GCG AAT GGA AGC ATT TTG AAG AAG GCA GAA      270
Ser Ala His Met Ala Gly Ala Asn Gly Ser Ile Leu Lys Lys Ala Glu
               30                  35                  40

GAG ACC TCT CGT GCA ATC ATG CAC AAA CCA GTG ATC TTC GGA GAA GAC      318
Glu Thr Ser Arg Ala Ile Met His Lys Pro Val Ile Phe Gly Glu Asp
           45                  50                  55

TAC ATT ACC GAG GCA GAC TTG CCT TAC ACA CCA CTC CAT TTA GAG GTC      366
Tyr Ile Thr Glu Ala Asp Leu Pro Tyr Thr Pro Leu His Leu Glu Val
       60                  65                  70

GAT GCT GAA ATG GAG CGG ATG TAT TAT CTT GGT CGT CGC GCG CTC ACC      414
Asp Ala Glu Met Glu Arg Met Tyr Tyr Leu Gly Arg Arg Ala Leu Thr
 75                  80                  85                  90

CAT GGC AAG AGA CGC AAA GTT TCT GTG AAT AAC AAG AGG AAC AGG AGA      462
His Gly Lys Arg Arg Lys Val Ser Val Asn Asn Lys Arg Asn Arg Arg
                95                 100                 105

AGG AAA GTG GCC AAA ACG TAC GTG GGG CGT GAT TCC ATT GTT GAG AAG      510
Arg Lys Val Ala Lys Thr Tyr Val Gly Arg Asp Ser Ile Val Glu Lys
           110                 115                 120

ATT GTA GTG CCC CAC ACC GAG AGA AAG GTT GAT ACC ACA GCA GCA GTG      558
Ile Val Val Pro His Thr Glu Arg Lys Val Asp Thr Thr Ala Ala Val
       125                 130                 135

GAA GAC ATT TGC AAT GAA GCT ACC ACT CAA CTT GTG CAT AAT AGT ATG      606
Glu Asp Ile Cys Asn Glu Ala Thr Thr Gln Leu Val His Asn Ser Met
   140                 145                 150

CCA AAG CGT AAG AAG CAG AAA AAC TTC TTG CCC GCC ACT TCA CTA AGT      654
Pro Lys Arg Lys Lys Gln Lys Asn Phe Leu Pro Ala Thr Ser Leu Ser
155                 160                 165                 170

AAC GTG TAT GCC CAA ACT TGG AGC ATA GTG CGC AAA CGC CAT ATG CAG      702
Asn Val Tyr Ala Gln Thr Trp Ser Ile Val Arg Lys Arg His Met Gln
               175                 180                 185

GTG GAG ATC ATT AGC AAG AAG AGC GTC CGA GCG AGG GTC AAG AGA TTT      750
Val Glu Ile Ile Ser Lys Lys Ser Val Arg Ala Arg Val Lys Arg Phe
           190                 195                 200

GAG GGC TCG GTG CAA TTG TTC GCA AGT GTG CGT CAC ATG TAT GGC GAG      798
Glu Gly Ser Val Gln Leu Phe Ala Ser Val Arg His Met Tyr Gly Glu
       205                 210                 215

AGG AAA AGG GTG GAC TTA CGT ATT GAC AAC TGG CAG CAA GAG ACA CTT      846
Arg Lys Arg Val Asp Leu Arg Ile Asp Asn Trp Gln Gln Glu Thr Leu
   220                 225                 230

CTA GAC CTT GCT AAA AGA TTT AAG AAT GAG AGA GTG GAT CAA TCG AAG      894
Leu Asp Leu Ala Lys Arg Phe Lys Asn Glu Arg Val Asp Gln Ser Lys
235                 240                 245                 250

CTC ACT TTT GGT TCA AGT GGC CTA GTT TTG AGG CAA GGC TCG TAC GGA      942
Leu Thr Phe Gly Ser Ser Gly Leu Val Leu Arg Gln Gly Ser Tyr Gly
               255                 260                 265

CCT GCG CAT TGG TAT CGA CAT GGT ATG TTC ATT GTA CGC GGT CGG TCG      990
Pro Ala His Trp Tyr Arg His Gly Met Phe Ile Val Arg Gly Arg Ser
           270                 275                 280

GAT GGG ATG TTG GTG GAT GCT CGT GCG AAG GTA ACG TTC GCT GTT TGT     1038
Asp Gly Met Leu Val Asp Ala Arg Ala Lys Val Thr Phe Ala Val Cys
       285                 290                 295
```

```
CAC TCA ATG ACA CAT TAT AGC GAC AAA TCA ATC TCT GAG GCA TTC TTC      1086
His Ser Met Thr His Tyr Ser Asp Lys Ser Ile Ser Glu Ala Phe Phe
    300                 305                 310

ATA CCA TAC TCT AAG AAA TTC TTG GAG TTG AGA CCA GAT GGA ATC TCC      1134
Ile Pro Tyr Ser Lys Lys Phe Leu Glu Leu Arg Pro Asp Gly Ile Ser
315                 320                 325                 330

CAT GAG TGT ACA AGA GGA GTA TCA GTT GAG CGG TGC GGT GAG GTG GCT      1182
His Glu Cys Thr Arg Gly Val Ser Val Glu Arg Cys Gly Glu Val Ala
                335                 340                 345

GCA ATC CTG ACA CAA GCA CTT TCA CCG TGT GGT AAG ATC ACA TGC AAA      1230
Ala Ile Leu Thr Gln Ala Leu Ser Pro Cys Gly Lys Ile Thr Cys Lys
            350                 355                 360

CGT TGC ATG GTT GAA ACA CCT GAC ATT GTT GAG GGT GAG TCG GGA GAA      1278
Arg Cys Met Val Glu Thr Pro Asp Ile Val Glu Gly Glu Ser Gly Glu
        365                 370                 375

AGT GTC ACC AAC CAA GGT AAG CTC CTA GCA ATG CTG AAA GAA CAG TAT      1326
Ser Val Thr Asn Gln Gly Lys Leu Leu Ala Met Leu Lys Glu Gln Tyr
    380                 385                 390

CCA GAT TTC CCA ATG GCC GAG AAA CTA CTC ACA AGG TTT TTG CAA CAG      1374
Pro Asp Phe Pro Met Ala Glu Lys Leu Leu Thr Arg Phe Leu Gln Gln
395                 400                 405                 410

AAA TCA CTA GTA AAT ACA AAT TTG ACA GCC TGC GTG AGC GTC AAA CAA      1422
Lys Ser Leu Val Asn Thr Asn Leu Thr Ala Cys Val Ser Val Lys Gln
                415                 420                 425

CTC ATT GGT GAC CGC AAA CAA GCT CCA TTC ACA CAC GTA CTG GCT GTC      1470
Leu Ile Gly Asp Arg Lys Gln Ala Pro Phe Thr His Val Leu Ala Val
            430                 435                 440

AGC GAA ATT CTG TTT AAA GGC AAT AAA CTA ACA GGG GCT GAT CTC GAA      1518
Ser Glu Ile Leu Phe Lys Gly Asn Lys Leu Thr Gly Ala Asp Leu Glu
        445                 450                 455

GAG GCA AGC ACA CAT ATG CTT GAA ATA GCA AGG TTC TTG AAC AAT CGC      1566
Glu Ala Ser Thr His Met Leu Glu Ile Ala Arg Phe Leu Asn Asn Arg
    460                 465                 470

ACT GAA AAT ATG CGC ATT GGC CAC CTT GGT TCT TTC AGA AAT AAA ATC      1614
Thr Glu Asn Met Arg Ile Gly His Leu Gly Ser Phe Arg Asn Lys Ile
475                 480                 485                 490

TCA TCG AAG GCC CAT GTG AAT AAC GCA CTC ATG TGT GAT AAT CAA CTT      1662
Ser Ser Lys Ala His Val Asn Asn Ala Leu Met Cys Asp Asn Gln Leu
                495                 500                 505

GAT CAG AAT GGG AAT TTT ATT TGG GGA CTA AGG GGT GCA CAC GCA AAG      1710
Asp Gln Asn Gly Asn Phe Ile Trp Gly Leu Arg Gly Ala His Ala Lys
            510                 515                 520

AGG TTT CTT AAA GGA TTT TTC ACT GAG ATT GAC CCA AAT GAA GGA TAC      1758
Arg Phe Leu Lys Gly Phe Phe Thr Glu Ile Asp Pro Asn Glu Gly Tyr
        525                 530                 535

GAT AAG TAT GTT ATC AGG AAA CAT ATC AGG GGT AGC AGA AAG CTA GCA      1806
Asp Lys Tyr Val Ile Arg Lys His Ile Arg Gly Ser Arg Lys Leu Ala
    540                 545                 550

ATT GGC AAT TTG ATA ATG TCA ACT GAC TTC CAG ACG CTC AGG CAA CAA      1854
Ile Gly Asn Leu Ile Met Ser Thr Asp Phe Gln Thr Leu Arg Gln Gln
555                 560                 565                 570
```

```
ATT CAA GGC GAA ACT ATT GAG CGT AAA GAA ATT GGG AAT CAC TGC ATT      1902
Ile Gln Gly Glu Thr Ile Glu Arg Lys Glu Ile Gly Asn His Cys Ile
                575                 580                 585

TCA ATG CGG AAT GGT AAT TAC GTG TAC CCA TGT TGT TGT GTT ACT CTT      1950
Ser Met Arg Asn Gly Asn Tyr Val Tyr Pro Cys Cys Cys Val Thr Leu
            590                 595                 600

GAA GAT GGT AAG GCT CAA TAT TCG GAT CTA AAG CAC CCA ACG AAG AGA      1998
Glu Asp Gly Lys Ala Gln Tyr Ser Asp Leu Lys His Pro Thr Lys Arg
        605                 610                 615

CAT CTG GTC ATT GGC AAC TCT GGC GAT TCA AAG TAC CTA GAC CTT CCA      2046
His Leu Val Ile Gly Asn Ser Gly Asp Ser Lys Tyr Leu Asp Leu Pro
    620                 625                 630

GTT CTC AAT GAA GAG AAA ATG TAT ATA GCT AAT GAA GGT TAT TGC TAC      2094
Val Leu Asn Glu Glu Lys Met Tyr Ile Ala Asn Glu Gly Tyr Cys Tyr
635                 640                 645                 650

ATG AAC ATT TTC TTT GCT CTA CTA GTG AAT GTC AAG GAA GAG GAT GCA      2142
Met Asn Ile Phe Phe Ala Leu Leu Val Asn Val Lys Glu Glu Asp Ala
                655                 660                 665

AAG GAC TTC ACC AAG TTT ATA AGG GAC ACA ATT GTT CCA AAG CTT GGA      2190
Lys Asp Phe Thr Lys Phe Ile Arg Asp Thr Ile Val Pro Lys Leu Gly
            670                 675                 680

GCG TGG CCA ACA ATG CAA GAT GTT GCA ACT GCA TGC TAC TTA CTT TCC      2238
Ala Trp Pro Thr Met Gln Asp Val Ala Thr Ala Cys Tyr Leu Leu Ser
        685                 690                 695

ATT CTT TAC CCA GAT GTC CTG AGA GCT GAA CTA CCC AGA ATT TTG GTT      2286
Ile Leu Tyr Pro Asp Val Leu Arg Ala Glu Leu Pro Arg Ile Leu Val
    700                 705                 710

GAT CAT GAC AAC AAA ACA ATG CAT GTT TTG GAT TCG TAT GGG TCT AGA      2334
Asp His Asp Asn Lys Thr Met His Val Leu Asp Ser Tyr Gly Ser Arg
715                 720                 725                 730

ACG ACA GGA TAC CAC ATG TTG AAA ATG AAC ACA ACA TCC CAG CTA ATT      2382
Thr Thr Gly Tyr His Met Leu Lys Met Asn Thr Thr Ser Gln Leu Ile
                735                 740                 745

GAA TTC GTT CAT TCA GGT TTG GAA TCC GAA ATG AAA ACT TAC AAT GTT      2430
Glu Phe Val His Ser Gly Leu Glu Ser Glu Met Lys Thr Tyr Asn Val
            750                 755                 760

GGA GGG ATG AAC CGA GAT GTG GTC ACA CAA GGT GCA ATT GAG ATG TTG      2478
Gly Gly Met Asn Arg Asp Val Val Thr Gln Gly Ala Ile Glu Met Leu
        765                 770                 775

ATC AAG TCT ATA TAC AAA CCA CAT CTC ATG AAG CAG TTA CTT GAG GAA      2526
Ile Lys Ser Ile Tyr Lys Pro His Leu Met Lys Gln Leu Leu Glu Glu
    780                 785                 790

GAG CCA TAC ATA ATT GTC CTG GCA ATA GTC TCC CCT TCA ATT TTA ATT      2574
Glu Pro Tyr Ile Ile Val Leu Ala Ile Val Ser Pro Ser Ile Leu Ile
795                 800                 805                 810

GCC ATG TAC AAC TCT GGA ACT TTT GAG CAG GCG TTA CAA ATG TGG TTG      2622
Ala Met Tyr Asn Ser Gly Thr Phe Glu Gln Ala Leu Gln Met Trp Leu
                815                 820                 825

CCA AAT ACA ATG AGG TTA GCT AAC CTC GCT GCC ATC TTG TCA GCC TTA      2670
Pro Asn Thr Met Arg Leu Ala Asn Leu Ala Ala Ile Leu Ser Ala Leu
            830                 835                 840
```

```
GCG CAA AAG TTA ACT TTG GCA GAT TTG TTC GTC CAG CAG CGT AAT TTG      2718
Ala Gln Lys Leu Thr Leu Ala Asp Leu Phe Val Gln Gln Arg Asn Leu
        845                 850                 855

ATT AAT GAG TAT GCG CAG GTA ATT TTG GAC AAT CTG ATT GAC GGT GTC      2766
Ile Asn Glu Tyr Ala Gln Val Ile Leu Asp Asn Leu Ile Asp Gly Val
    860                 865                 870

AGG GTT AAT CAT TCG CTA TCC CTA GCA ATG GAA ATT GTT ACT ATT AAG      2814
Arg Val Asn His Ser Leu Ser Leu Ala Met Glu Ile Val Thr Ile Lys
875                 880                 885                 890

CTG GCC ACC CAA GAG ATG GAC ATG GCG TTG AGG GAA GGT GGC TAT GCT      2862
Leu Ala Thr Gln Glu Met Asp Met Ala Leu Arg Glu Gly Gly Tyr Ala
                895                 900                 905

GTG ACC TCT GAA AAG GTG CAT GAA ATG TTG GAA AAA AAC TAT GTA AAG      2910
Val Thr Ser Glu Lys Val His Glu Met Leu Glu Lys Asn Tyr Val Lys
            910                 915                 920

GCT TTG AAG GAT GCA TGG GAC GAA TTA ACT TGG TTG GAA AAA TTC TCC      2958
Ala Leu Lys Asp Ala Trp Asp Glu Leu Thr Trp Leu Glu Lys Phe Ser
        925                 930                 935

GCA ATC AGG CAT TCA AGA AAG CTC TTG AAA TTT GGG CGA AAG CCT TTA      3006
Ala Ile Arg His Ser Arg Lys Leu Leu Lys Phe Gly Arg Lys Pro Leu
    940                 945                 950

ATC ATG AAA AAC ACC GTA GAT TGC GGC GGA CAT ATA GAC TTG TCT GTG      3054
Ile Met Lys Asn Thr Val Asp Cys Gly Gly His Ile Asp Leu Ser Val
955                 960                 965                 970

AAA TCG CTT TTC AAG TTC CAC TTG GAA CTC CTG AAG GGA ACC ATC TCA      3102
Lys Ser Leu Phe Lys Phe His Leu Glu Leu Leu Lys Gly Thr Ile Ser
                975                 980                 985

AGA GCC GTA AAT GGT GGC GCA AGA AAG GTA AGA GTA GCG AAG AAT GCC      3150
Arg Ala Val Asn Gly Gly Ala Arg Lys Val Arg Val Ala Lys Asn Ala
            990                 995                1000

ATG ACA AAA GGG GTT TTT CTC AAA ATC TAC AGC ATG CTT CCT GAC GTC      3198
Met Thr Lys Gly Val Phe Leu Lys Ile Tyr Ser Met Leu Pro Asp Val
        1005                1010                1015

TAC AAG TTT ATC ACA GTC TCG AGT GTC CTT TCC TTG TTG TTG ACA TTC      3246
Tyr Lys Phe Ile Thr Val Ser Ser Val Leu Ser Leu Leu Leu Thr Phe
    1020                1025                1030

TTA TTT CAA ATT GAC TGC ATG ATA AGG GCA CAC CGA GAG GCG AAG GTT      3294
Leu Phe Gln Ile Asp Cys Met Ile Arg Ala His Arg Glu Ala Lys Val
1035                1040                1045                1050

GCT GCA CAG TTG CAG AAA GAG AGC GAG TGG GAC AAT ATC ATC AAT AGA      3342
Ala Ala Gln Leu Gln Lys Glu Ser Glu Trp Asp Asn Ile Ile Asn Arg
                1055                1060                1065

ACT TTC CAG TAT TCT AAG CTT GAA AAT CCT ATT GGC TAT CGC TCT ACA      3390
Thr Phe Gln Tyr Ser Lys Leu Glu Asn Pro Ile Gly Tyr Arg Ser Thr
            1070                1075                1080

GCG GAG GAA AGA CTC CAA TCA GAA CAC CCC GAG GCT TTC GAG TAC TAC      3438
Ala Glu Glu Arg Leu Gln Ser Glu His Pro Glu Ala Phe Glu Tyr Tyr
        1085                1090                1095

AAG TTT TGC ATT GGA AAG GAA GAC CTC GTT GAA CAG GCA AAA CAA CCG      3486
Lys Phe Cys Ile Gly Lys Glu Asp Leu Val Glu Gln Ala Lys Gln Pro
    1100                1105                1110
```

```
GAG ATA GCA TAC TTT GAA AAG ATT ATA GCT TTC ATC ACA CTT GTA TTA      3534
Glu Ile Ala Tyr Phe Glu Lys Ile Ile Ala Phe Ile Thr Leu Val Leu
1115                1120                1125                1130

ATG GCT TTT GAC GCT GAG CGG AGT GAT GGA GTG TTC AAG ATA CTC AAT      3582
Met Ala Phe Asp Ala Glu Arg Ser Asp Gly Val Phe Lys Ile Leu Asn
                    1135                1140                1145

AAG TTC AAA GGA ATA CTG AGC TCA ACG GAG AGG GAG ATC ATC TAC ACG      3630
Lys Phe Lys Gly Ile Leu Ser Ser Thr Glu Arg Glu Ile Ile Tyr Thr
            1150                1155                1160

CAG AGT TTG GAT GAT TAC GTT ACA ACC TTT GAT GAC AAT ATG ACA ATC      3678
Gln Ser Leu Asp Asp Tyr Val Thr Thr Phe Asp Asp Asn Met Thr Ile
        1165                1170                1175

AAC CTC GAG TTG AAT ATG GAT GAA CTC CAC AAG ACG AGC CTT CCT GGA      3726
Asn Leu Glu Leu Asn Met Asp Glu Leu His Lys Thr Ser Leu Pro Gly
    1180                1185                1190

GTC ACT TTT AAG CAA TGG TGG AAC AAC CAA ATC AGC CGA GGC AAC GTG      3774
Val Thr Phe Lys Gln Trp Trp Asn Asn Gln Ile Ser Arg Gly Asn Val
1195                1200                1205                1210

AAG CCA CAT TAT AGA ACT GAG GGG CAC TTC ATG GAG TTT ACC AGA GAT      3822
Lys Pro His Tyr Arg Thr Glu Gly His Phe Met Glu Phe Thr Arg Asp
                    1215                1220                1225

ACT GCG GCA TCG GTT GCC AGC GAG ATA TCA CAC TCA CCC GCA AGA GAT      3870
Thr Ala Ala Ser Val Ala Ser Glu Ile Ser His Ser Pro Ala Arg Asp
            1230                1235                1240

TTT CTT GTG AGA GGT GCT GTT GGA TCT GGA AAA TCC ACA GGA CTT CCA      3918
Phe Leu Val Arg Gly Ala Val Gly Ser Gly Lys Ser Thr Gly Leu Pro
        1245                1250                1255

TAC CAT TTA TCA AAG AGA GGG AGA GTG TTA ATG CTT GAG CCT ACC AGA      3966
Tyr His Leu Ser Lys Arg Gly Arg Val Leu Met Leu Glu Pro Thr Arg
    1260                1265                1270

CCA CTC ACA GAT AAC ATG CAC AAG CAA CTG AGA AGT GAA CCA TTT AAC      4014
Pro Leu Thr Asp Asn Met His Lys Gln Leu Arg Ser Glu Pro Phe Asn
1275                1280                1285                1290

TGC TTC CCA ACT TTG AGG ATG AGA GGG AAG TCA ACT TTT GGG TCA TCA      4062
Cys Phe Pro Thr Leu Arg Met Arg Gly Lys Ser Thr Phe Gly Ser Ser
                    1295                1300                1305

CCG ATC ACA GTC ATG ACT AGT GGA TTC GCT TTA CAC CAC TTT GCA CGA      4110
Pro Ile Thr Val Met Thr Ser Gly Phe Ala Leu His His Phe Ala Arg
            1310                1315                1320

AAC ATA GCT GAG GTA AAA ACA TAC GAT TTT GTC ATA ATT GAT GAA TGT      4158
Asn Ile Ala Glu Val Lys Thr Tyr Asp Phe Val Ile Ile Asp Glu Cys
        1325                1330                1335

CAT GTG AAT GAT GCT TCT GCT ATA GCG TTT AGG AAT CTA CTG TTT GAA      4206
His Val Asn Asp Ala Ser Ala Ile Ala Phe Arg Asn Leu Leu Phe Glu
    1340                1345                1350

CAT GAA TTT GAA GGA AAA GTC CTC AAA GTG TCA GCC ACA CCA CCA GGT      4254
His Glu Phe Glu Gly Lys Val Leu Lys Val Ser Ala Thr Pro Pro Gly
1355                1360                1365                1370

AGA GAA GTT GAA TTT ACA ACT CAG TTT CCC GTG AAA CTC AAG ATA GAA      4302
Arg Glu Val Glu Phe Thr Thr Gln Phe Pro Val Lys Leu Lys Ile Glu
                    1375                1380                1385
```

```
GAG GCT CTT AGC TTT CAG GAA TTT GTA AGT TTA CAA GGG ACA GGT GCC     4350
Glu Ala Leu Ser Phe Gln Glu Phe Val Ser Leu Gln Gly Thr Gly Ala
            1390                1395                1400

AAC GCC GAT GTG ATT AGT TGT GGC GAC AAC ATA CTA GTA TAT GTT GCT     4398
Asn Ala Asp Val Ile Ser Cys Gly Asp Asn Ile Leu Val Tyr Val Ala
        1405                1410                1415

AGC TAC AAT GAT GTT GAT AGT CTT GGC AAG CTC CTT GTG CAA AAG GGA     4446
Ser Tyr Asn Asp Val Asp Ser Leu Gly Lys Leu Leu Val Gln Lys Gly
    1420                1425                1430

TAC AAA GTG TCG AAG ATT GAT GGA AGA ACA ATG AAG AGT GGA GGA ACT     4494
Tyr Lys Val Ser Lys Ile Asp Gly Arg Thr Met Lys Ser Gly Gly Thr
1435                1440                1445                1450

GAA ATA ATC ACT GAA GGT ACT TCA GTG AAA AAG CAT TTC ATA GTC GCA     4542
Glu Ile Ile Thr Glu Gly Thr Ser Val Lys Lys His Phe Ile Val Ala
                1455                1460                1465

ACT AAC ATT ATT GAG AAT GGT GTA ACC ATT GAC ATT GAT GTA GTT GTG     4590
Thr Asn Ile Ile Glu Asn Gly Val Thr Ile Asp Ile Asp Val Val Val
            1470                1475                1480

GAT TTT GGG ACT AAG GTT GTA CCA GTT TTG GAT GTG GAC AAT AGA GCG     4638
Asp Phe Gly Thr Lys Val Val Pro Val Leu Asp Val Asp Asn Arg Ala
        1481                1490                1495

GTG CAG TAC AAC AAA ACT GTG GTG AGT TAT GGG GAG CGC ATC CAA AAA     4686
Val Gln Tyr Asn Lys Thr Val Val Ser Tyr Gly Glu Arg Ile Gln Lys
    1500                1505                1510

CTC GGT AGA GTT GGG CGA CAC AAG GAA GGA GTA GCA CTT CGA ATT GGC     4734
Leu Gly Arg Val Gly Arg His Lys Glu Gly Val Ala Leu Arg Ile Gly
1515                1520                1525                1530

CAA ACA AAT AAA ACA CTG GTT GAA ATT CCA GAA ATG GTT GCC ACT GAA     4782
Gln Thr Asn Lys Thr Leu Val Glu Ile Pro Glu Met Val Ala Thr Glu
                1535                1540                1545

GCT GCC TTT CTA TGC TTC ATG TAC AAT TTG CCA GTG ACA ACA CAG AGT     4830
Ala Ala Phe Leu Cys Phe Met Tyr Asn Leu Pro Val Thr Thr Gln Ser
            1550                1555                1560

GTT TCA ACC ACA CTG CTG GAA AAT GCC ACA TTA TTA CAA GCT AGA ACT     4878
Val Ser Thr Thr Leu Leu Glu Asn Ala Thr Leu Leu Gln Ala Arg Thr
        1565                1570                1575

ATG GCA CAG TTT GAG CTA TCA TAT TTT TAC ACA ATT AAT TTT GTG CGA     4926
Met Ala Gln Phe Glu Leu Ser Tyr Phe Tyr Thr Ile Asn Phe Val Arg
    1580                1585                1590

TTT GAT GGT AGT ATG CAT CCA GTC ATA CAT GAC AAG CTG AAG CGC TTT     4974
Phe Asp Gly Ser Met His Pro Val Ile His Asp Lys Leu Lys Arg Phe
1595                1600                1605                1610

AAG CTA CAC ACT TGT GAG ACA TTC CTC AAT AAG TTG GCG ATC CCA AAT     5022
Lys Leu His Thr Cys Glu Thr Phe Leu Asn Lys Leu Ala Ile Pro Asn
                1615                1620                1625

AAA GGC TTA TCC TCT TGG CTT ACG AGT GGA GAG TAT AAG CGA CTT GGT     5070
Lys Gly Leu Ser Ser Trp Leu Thr Ser Gly Glu Tyr Lys Arg Leu Gly
            1630                1635                1640

TAC ATA GCA GAG GAT GCT GGC ATA AGA ATC CCA TTC GTG TGC AAA GAA     5118
Tyr Ile Ala Glu Asp Ala Gly Ile Arg Ile Pro Phe Val Cys Lys Glu
        1645                1650                1655
```

```
ATT CCA GAC TCC TTG CAT GAG GAA ATT TGG CAC ATT GTA GTC GCC CAT      5166
Ile Pro Asp Ser Leu His Glu Glu Ile Trp His Ile Val Val Ala His
    1660                1665                1670

AAA GGT GAC TCG GGT ATT GGG AGG CTC ACT AGC GTA CAG GCA GCA AAG      5214
Lys Gly Asp Ser Gly Ile Gly Arg Leu Thr Ser Val Gln Ala Ala Lys
1675                1680                1685                1690

GTT GTT TAT ACT CTG CAA ACG GAT GTG CAC TCA ATT GCG AGG ACT CTA      5262
Val Val Tyr Thr Leu Gln Thr Asp Val His Ser Ile Ala Arg Thr Leu
                1695                1700                1705

GCA TGC ATC AAT AGA CGC ATA GCA GAT GAA CAA ATG AAG CAG AGT CAT      5310
Ala Cys Ile Asn Arg Arg Ile Ala Asp Glu Gln Met Lys Gln Ser His
            1710                1715                1720

TTT GAA GCC GCA ACT GGG AGA GCA TTT TCC TTC ACA AAT TAC TCA ATA      5358
Phe Glu Ala Ala Thr Gly Arg Ala Phe Ser Phe Thr Asn Tyr Ser Ile
        1725                1730                1735

CAA AGC ATA TTT GAC ACG CTG AAA GCA AAT TAT GCT ACA AAG CAT ACG      5406
Gln Ser Ile Phe Asp Thr Leu Lys Ala Asn Tyr Ala Thr Lys His Thr
    1740                1745                1750

AAA GAA AAT ATT GCA GTG CTT CAG CAG GCA AAA GAT CAA TTG CTA GAG      5454
Lys Glu Asn Ile Ala Val Leu Gln Gln Ala Lys Asp Gln Leu Leu Glu
1755                1760                1765                1770

TTT TCG AAC CTA GCA AAG GAT CAA GAT GTC ACG GGT ATC ATC CAA GAC      5502
Phe Ser Asn Leu Ala Lys Asp Gln Asp Val Thr Gly Ile Ile Gln Asp
                1775                1780                1785

TTC AAT CAC CTG GAA ACT ATC TAT CTC CAA TCA GAT AGC GAA GTG GCT      5550
Phe Asn His Leu Glu Thr Ile Tyr Leu Gln Ser Asp Ser Glu Val Ala
            1790                1795                1800

AAG CAT CTG AAG CTT AAA AGT CAC TGG AAT AAA AGC CAA ATC ACT AGG      5598
Lys His Leu Lys Leu Lys Ser His Trp Asn Lys Ser Gln Ile Thr Arg
        1805                1810                1815

GAC ATC ATA ATA GCT TTG TCT GTG TTA ATT GGT GGT GGA TGG ATG CTT      5646
Asp Ile Ile Ile Ala Leu Ser Val Leu Ile Gly Gly Gly Trp Met Leu
    1820                1825                1830

GCA ACG TAC TTC AAG GAC AAG TTC AAT GAA CCA GTC TAT TTC CAA GGG      5694
Ala Thr Tyr Phe Lys Asp Lys Phe Asn Glu Pro Val Tyr Phe Gln Gly
1835                1840                1845                1850

AAG AAG AAT CAG AAG CAC AAG CTT AAG ATG AGA GAG GCG CGT GGG GCT      5742
Lys Lys Asn Gln Lys His Lys Leu Lys Met Arg Glu Ala Arg Gly Ala
                1855                1860                1865

AGA GGG CAA TAT GAG GTT GCA GCG GAG CCA GAG GCG CTA GAA CAT TAC      5790
Arg Gly Gln Tyr Glu Val Ala Ala Glu Pro Glu Ala Leu Glu His Tyr
            1870                1875                1880

TTT GGA AGC GCA TAT AAT AAC AAA GGA AAG CGC AAG GGC ACC ACG AGA      5838
Phe Gly Ser Ala Tyr Asn Asn Lys Gly Lys Arg Lys Gly Thr Thr Arg
        1885                1890                1895

GGA ATG GGT GCA AAG TCT CGG AAA TTC ATA AAC ATG TAT GGG TTT GAT      5886
Gly Met Gly Ala Lys Ser Arg Lys Phe Ile Asn Met Tyr Gly Phe Asp
    1900                1905                1910

CCA ACT GAT TTT TCA TAC ATT AGG TTT GTG GAT CCA TTG ACA GGT CAC      5934
Pro Thr Asp Phe Ser Tyr Ile Arg Phe Val Asp Pro Leu Thr Gly His
1915                1920                1925                1930
```

```
ACT ATT GAT GAG TCC ACA AAC GCA CCT ATT GAT TTA GTG CAG CAT GAG      5982
Thr Ile Asp Glu Ser Thr Asn Ala Pro Ile Asp Leu Val Gln His Glu
                1935                1940                1945

TTT GGA AAG GTT AGA ACA CGC ATG TTA ATT GAC GAT GAG ATA GAG CCT      6030
Phe Gly Lys Val Arg Thr Arg Met Leu Ile Asp Asp Glu Ile Glu Pro
            1950                1955                1960

CAA AGT CTT AGC ACC CAC ACC ACA ATC CAT GCT TAT TTG GTG AAT AGT      6078
Gln Ser Leu Ser Thr His Thr Thr Ile His Ala Tyr Leu Val Asn Ser
        1965                1970                1975

GGC ACG AAG AAA GTT CTT AAG GTT GAT TTA ACA CCA CAC TCG TCG CTA      6126
Gly Thr Lys Lys Val Leu Lys Val Asp Leu Thr Pro His Ser Ser Leu
    1980                1985                1990

CGT GCG AGT GAG AAA TCA ACA GCA ATA ATG GGA TTT CCT GAA AGG GAG      6174
Arg Ala Ser Glu Lys Ser Thr Ala Ile Met Gly Phe Pro Glu Arg Glu
1995                2000                2005                2010

AAT GAA TTG CGT CAA ACC GGC ATG GCA GTG CCA GTG GCT TAT GAT CAA      6222
Asn Glu Leu Arg Gln Thr Gly Met Ala Val Pro Val Ala Tyr Asp Gln
                2015                2020                2025

TTG CCA CCA AAG AAT GAG GAC TTG ACG TTT GAA GGA GAA AGC TTG TTT      6270
Leu Pro Pro Lys Asn Glu Asp Leu Thr Phe Glu Gly Glu Ser Leu Phe
            2030                2035                2040

AAG GGA CCA CGT GAT TAC AAC CCG ATA TCG AGC ACC ATT TGT CAT TTG      6318
Lys Gly Pro Arg Asp Tyr Asn Pro Ile Ser Ser Thr Ile Cys His Leu
        2045                2050                2055

ACG AAT GAA TCT GAT GGG CAC ACA ACA TCG TTG TAT GGT ATT GGA TTT      6366
Thr Asn Glu Ser Asp Gly His Thr Thr Ser Leu Tyr Gly Ile Gly Phe
    2060                2065                2070

GGT CCC TTC ATC ATT ACA AAC AAG CAC TTG TTT AGA AGA AAT AAT GGA      6414
Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe Arg Arg Asn Asn Gly
2075                2080                2085                2090

ACA CTG TTG GTC CAA TCA CTA CAT GGT GTA TTC AAG GTC AAG AAC ACC      6462
Thr Leu Leu Val Gln Ser Leu His Gly Val Phe Lys Val Lys Asn Thr
                2095                2100                2105

ACG ACT TTG CAA CAA CAC CTC ATT GAT GGG AGG GAC ATG ATA ATT ATT      6510
Thr Thr Leu Gln Gln His Leu Ile Asp Gly Arg Asp Met Ile Ile Ile
            2110                2115                2120

CGC ATG CCT AAG GAT TTC CCA CCA TTT CCT CAA AAG CTG AAA TTT AGA      6558
Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln Lys Leu Lys Phe Arg
        2125                2130                2135

GAG CCA CAA AGG GAA GAG CGC ATA TGT CTT GTG ACA ACC AAC TTC CAA      6606
Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val Thr Thr Asn Phe Gln
    2140                2145                2150

ACT AAG AGC ATG TCT AGC ATG GTG TCA GAC ACT AGT TGC ACA TTC CCT      6654
Thr Lys Ser Met Ser Ser Met Val Ser Asp Thr Ser Cys Thr Phe Pro
2155                2160                2165                2170

TCA TCT GAT GGC ATA TTC TGG AAG CAT TGG ATT CAA ACC AAG GAT GGG      6702
Ser Ser Asp Gly Ile Phe Trp Lys His Trp Ile Gln Thr Lys Asp Gly
                2175                2180                2185

CAG TGT GGC AGT CCA TTA GTA TCA ACT AGA GAT GGG TTC ATT GTT GGT      6750
Gln Cys Gly Ser Pro Leu Val Ser Thr Arg Asp Gly Phe Ile Val Gly
            2190                2195                2200
```

```
ATA CAC TCA GCA TCG AAT TTC ACC AAC ACA AAC AAT TAT TTC ACA AGC      6798
Ile His Ser Ala Ser Asn Phe Thr Asn Thr Asn Asn Tyr Phe Thr Ser
        2205                2210                2215

GTG CCG AAA AAC TTC ATG GAA TTG TTG ACA AAT CAG GAG GCG CAG CAG      6846
Val Pro Lys Asn Phe Met Glu Leu Leu Thr Asn Gln Glu Ala Gln Gln
    2220                2225                2230

TGG GTT AGT GGT TGG CGA TTA AAT GCT GAC TCA GTA TTG TGG GGG GGC      6894
Trp Val Ser Gly Trp Arg Leu Asn Ala Asp Ser Val Leu Trp Gly Gly
2235                2240                2245                2250

CAT AAA GTT TTC ATG AGC AAA CCT GAA GAG CCT TTT CAG CCA GTT AAG      6942
His Lys Val Phe Met Ser Lys Pro Glu Glu Pro Phe Gln Pro Val Lys
                2255                2260                2265

GAA GCG ACT CAA CTC ATG AAT GAA TTG GTG TAC TCG CAA GGG GAG AAG      6990
Glu Ala Thr Gln Leu Met Asn Glu Leu Val Tyr Ser Gln Gly Glu Lys
            2270                2275                2280

AGG AAA TGG GTC GTG GAA GCA CTG TCA GGG AAC TTG AGG CCA GTG GCT      7038
Arg Lys Trp Val Val Glu Ala Leu Ser Gly Asn Leu Arg Pro Val Ala
        2285                2290                2295

GAG TGT CCC AGT CAG TTA GTC ACA AAG CAT GTG GTT AAA GGA AAG TGT      7086
Glu Cys Pro Ser Gln Leu Val Thr Lys His Val Val Lys Gly Lys Cys
    2300                2305                2310

CCC CTC TTT GAG CTC TAC TTG CAG TTG AAT CCA GAA AAG GAA GCA TAT      7134
Pro Leu Phe Glu Leu Tyr Leu Gln Leu Asn Pro Glu Lys Glu Ala Tyr
2315                2320                2325                2330

TTT AAA CCG ATG ATG GGA GCA TAT AAG CCA AGT CGA CTT AAT AGA GAG      7182
Phe Lys Pro Met Met Gly Ala Tyr Lys Pro Ser Arg Leu Asn Arg Glu
                2335                2340                2345

GCG TTC CTC AAG GAC ATT CTA AAA TAT GCT AGT GAA ATT GAG ATT GGG      7230
Ala Phe Leu Lys Asp Ile Leu Lys Tyr Ala Ser Glu Ile Glu Ile Gly
            2350                2355                2360

AAT GTG GAT TGT GAC TTG CTG GAG CTT GCA ATA AGC ATG CTC GTC ACA      7278
Asn Val Asp Cys Asp Leu Leu Glu Leu Ala Ile Ser Met Leu Val Thr
        2365                2370                2375

AAG CTC AAG GCG TTA GGA TTC CCA ACT GTG AAC TAC ATC ACT GAC CCA      7326
Lys Leu Lys Ala Leu Gly Phe Pro Thr Val Asn Tyr Ile Thr Asp Pro
    2380                2385                2390

GAG GAA ATT TTT AGT GCA TTG AAT ATG AAA GCA GCT ATG GGA GCA CTA      7374
Glu Glu Ile Phe Ser Ala Leu Asn Met Lys Ala Ala Met Gly Ala Leu
2395                2400                2405                2410

TAC AAA GGC AAG AAG AAA GAA GCT CTC AGC GAG CTC ACA CTA GAT GAG      7422
Tyr Lys Gly Lys Lys Lys Glu Ala Leu Ser Glu Leu Thr Leu Asp Glu
                2415                2420                2425

CAG GAG GCA ATG CTC AAA GCA AGT TGC CTG CGA CTG TAT ACG GGA AAG      7470
Gln Glu Ala Met Leu Lys Ala Ser Cys Leu Arg Leu Tyr Thr Gly Lys
            2430                2435                2440

TTG GGA ATT TGG AAT GGC TCA TTG AAA GCA GAG TTG CGT CCA ATT GAG      7518
Leu Gly Ile Trp Asn Gly Ser Leu Lys Ala Glu Leu Arg Pro Ile Glu
        2445                2450                2455

AAG GTT GAA AAC AAC AAA ACG CGA ACT TTC ACA GCA GCA CCA ATA GAC      7566
Lys Val Glu Asn Asn Lys Thr Arg Thr Phe Thr Ala Ala Pro Ile Asp
    2460                2465                2470
```

```
ACT CTT CTT GCT GGT AAA GTT TGC GTG GAT GAT TTC AAC AAT CAA TTT      7614
Thr Leu Leu Ala Gly Lys Val Cys Val Asp Asp Phe Asn Asn Gln Phe
2475                2480                2485                2490

TAT GAT CTC AAC ATA AAG GCA CCA TGG ACA GTT GGT ATG ACT AAG TTT      7662
Tyr Asp Leu Asn Ile Lys Ala Pro Trp Thr Val Gly Met Thr Lys Phe
                2495                2500                2505

TAT CAG GGG TGG AAT GAA TTG ATG GAG GCT TTA CCA AGT GGG TGG GTG      7710
Tyr Gln Gly Trp Asn Glu Leu Met Glu Ala Leu Pro Ser Gly Trp Val
            2510                2515                2520

TAT TGT GAC GCT GAT GGT TCG CAA TTC GAC AGT TCC TTG ACT CCA TTC      7758
Tyr Cys Asp Ala Asp Gly Ser Gln Phe Asp Ser Ser Leu Thr Pro Phe
        2525                2530                2535

CTC ATT AAT GCT GTA TTG AAA GTG CGA CTT GCC TTC ATG GAG GAA TGG      7806
Leu Ile Asn Ala Val Leu Lys Val Arg Leu Ala Phe Met Glu Glu Trp
    2540                2545                2550

GAT ATT GGT GAG CAA ATG CTG CGA AAT TTG TAC ACT GAG ATA GTG TAT      7854
Asp Ile Gly Glu Gln Met Leu Arg Asn Leu Tyr Thr Glu Ile Val Tyr
2555                2560                2565                2570

ACA CCA ATC CTC ACA CCG GAT GGT ACT ATC ATT AAG AAG CAT AAA GGC      7902
Thr Pro Ile Leu Thr Pro Asp Gly Thr Ile Ile Lys Lys His Lys Gly
                2575                2580                2585

AAC AAT AGC GGG CAA CCT TCA ACA GTG GTG GAC AAC ACA CTC ATG GTC      7950
Asn Asn Ser Gly Gln Pro Ser Thr Val Val Asp Asn Thr Leu Met Val
            2590                2595                2600

ATT ATT GCA ATG TTA TAC ACA TGT GAG AAG TGT GGA ATC AAC AAG GAA      7998
Ile Ile Ala Met Leu Tyr Thr Cys Glu Lys Cys Gly Ile Asn Lys Glu
        2605                2610                2615

GAG ATT GTG TAT TAC GTC AAT GGC GAT GAC CTA TTG ATT GCC ATT CAC      8046
Glu Ile Val Tyr Tyr Val Asn Gly Asp Asp Leu Leu Ile Ala Ile His
    2620                2625                2630

CCA GAT AAA GCT GAG AGG TTG AGT AGA TTC AAA GAA TCT TTC GGA GAG      8094
Pro Asp Lys Ala Glu Arg Leu Ser Arg Phe Lys Glu Ser Phe Gly Glu
2635                2640                2645                2650

TTG GGC CTG AAA TAT GAA TTT GAC TGT ACC ACC AGG GAC AAG ACA CAG      8142
Leu Gly Leu Lys Tyr Glu Phe Asp Cys Thr Thr Arg Asp Lys Thr Gln
                2655                2660                2665

TTG TGG TTC ATG TCA CAC AGG GCT TTG GAG AGG GAT GGC ATG TAT ATA      8190
Leu Trp Phe Met Ser His Arg Ala Leu Glu Arg Asp Gly Met Tyr Ile
            2670                2675                2680

CCA AAG CTA GAA GAA GAA AGG ATT GTT TCT ATT TTG GAA TGG GAC AGA      8238
Pro Lys Leu Glu Glu Glu Arg Ile Val Ser Ile Leu Glu Trp Asp Arg
        2685                2690                2695

TCC AAA GAG CCG TCA CAT AGG CTT GAA GCC ATC TGT GCA TCA ATG ATT      8286
Ser Lys Glu Pro Ser His Arg Leu Glu Ala Ile Cys Ala Ser Met Ile
    2700                2705                2710

GAA GCA TGG GGT TAT GAC AAG CTG GTT GAA GAA ATC CGC AAT TTC TAT      8334
Glu Ala Trp Gly Tyr Asp Lys Leu Val Glu Glu Ile Arg Asn Phe Tyr
2715                2720                2725                2730

GCA TGG GTT TTG GAA CAA GCG CCG TAT TCA CAG CTT GCA GAA GAA GGA      8382
Ala Trp Val Leu Glu Gln Ala Pro Tyr Ser Gln Leu Ala Glu Glu Gly
                2735                2740                2745
```

```
AAG GCG CCA TAT CTG GCT GAG ACT GCG CTT AAG TTT TTG TAC ACA TCT     8430
Lys Ala Pro Tyr Leu Ala Glu Thr Ala Leu Lys Phe Leu Tyr Thr Ser
            2750                2755                2760

CAG CAC GGA ACA AAC TCT GAG ATA GAA GAG TAT TTA AAA GTG TTG TAT     8478
Gln His Gly Thr Asn Ser Glu Ile Glu Glu Tyr Leu Lys Val Leu Tyr
        2765                2770                2775

GAT TAC GAT ATT CCA ACG ACT GAG AAT CTT TAT TTT CAG AGT GGC ACT     8526
Asp Tyr Asp Ile Pro Thr Thr Glu Asn Leu Tyr Phe Gln Ser Gly Thr
    2780                2785                2790

GTG GAT GCT GGT GCT GAC GCT GGT AAG AAG AAA GAT CAA AAG GAT GAT     8574
Val Asp Ala Gly Ala Asp Ala Gly Lys Lys Lys Asp Gln Lys Asp Asp
2795                2800                2805                2810

AAA GTC GCT GAG CAG GCT TCA AAG GAT AGG GAT GTT AAT GCT GGA ACT     8622
Lys Val Ala Glu Gln Ala Ser Lys Asp Arg Asp Val Asn Ala Gly Thr
            2815                2820                2825

TCA GGA ACA TTC TCA GTT CCA CGA ATA AAT GCT ATG GCC ACA AAA CTT     8670
Ser Gly Thr Phe Ser Val Pro Arg Ile Asn Ala Met Ala Thr Lys Leu
            2830                2835                2840

CAA TAT CCA AGG ATG AGG GGA GAG GTG GTT GTA AAC TTG AAT CAC CTT     8718
Gln Tyr Pro Arg Met Arg Gly Glu Val Val Val Asn Leu Asn His Leu
        2845                2850                2855

TTA GGA TAC AAG CCA CAG CAA ATT GAT TTG TCA AAT GCT CGA GCC ACA     8766
Leu Gly Tyr Lys Pro Gln Gln Ile Asp Leu Ser Asn Ala Arg Ala Thr
    2860                2865                2870

CAT GAG CAG TTT GCC GCG TGG CAT CAG GCA GTG ATG ACA GCC TAT GGA     8814
His Glu Gln Phe Ala Ala Trp His Gln Ala Val Met Thr Ala Tyr Gly
2875                2880                2885                2890

GTG AAT GAA GAG CAA ATG AAA ATA TTG CTA AAT GGA TTT ATG GTG TGG     8862
Val Asn Glu Glu Gln Met Lys Ile Leu Leu Asn Gly Phe Met Val Trp
            2895                2900                2905

TGC ATA GAA AAT GGG ACT TCC CCA AAT TTG AAC GGA ACT TGG GTT ATG     8910
Cys Ile Glu Asn Gly Thr Ser Pro Asn Leu Asn Gly Thr Trp Val Met
            2910                2915                2920

ATG GAT GGT GAG GAT CAA GTT TCA TAC CCG CTG AAA CCA ATG GTT GAA     8958
Met Asp Gly Glu Asp Gln Val Ser Tyr Pro Leu Lys Pro Met Val Glu
        2925                2930                2935

AAC GCG CAG CCA ACA CTG AGG CAA ATT ATG ACA CAC TTC AGT GAC CTG     9006
Asn Ala Gln Pro Thr Leu Arg Gln Ile Met Thr His Phe Ser Asp Leu
    2940                2945                2950

GCT GAA GCG TAT ATT GAG ATG AGG AAT AGG GAG CGA CCA TAC ATG CCT     9054
Ala Glu Ala Tyr Ile Glu Met Arg Asn Arg Glu Arg Pro Tyr Met Pro
2955                2960                2965                2970

AGG TAT GGT CTA CAG AGA AAC ATT ACA GAC ATG AGT TTG TCA CGC TAT     9102
Arg Tyr Gly Leu Gln Arg Asn Ile Thr Asp Met Ser Leu Ser Arg Tyr
            2975                2980                2985

GCG TTC GAC TTC TAT GAG CTA ACT TCA AAA ACA CCT GTT AGA GCG AGG     9150
Ala Phe Asp Phe Tyr Glu Leu Thr Ser Lys Thr Pro Val Arg Ala Arg
            2990                2995                3000

GAG GCG CAT ATG CAA ATG AAA GCT GCT GCA GTA CGA AAC AGT GGA ACT     9198
Glu Ala His Met Gln Met Lys Ala Ala Ala Val Arg Asn Ser Gly Thr
        3005                3010                3015
```

```
AGG TTA TTT GGT CTT GAT GGC AAC GTG GGT ACT GCA GAG GAA GAC ACT      9246
Arg Leu Phe Gly Leu Asp Gly Asn Val Gly Thr Ala Glu Glu Asp Thr
    3020                3025                3030

GAA CGG CAC ACA GCG CAC GAT GTG AAC CGT AAC ATG CAC ACA CTA TTA      9294
Glu Arg His Thr Ala His Asp Val Asn Arg Asn Met His Thr Leu Leu
3035                3040                3045                3050

GGG GTC CGC CAG TGA TAGTTTCTGC GTGTCTTTGC TTTCCGCTTT TAAGCTTATT     9349
Gly Val Arg Gln

GTAATATATA TGAATAGCTA TTCACAGTGG GACTTGGTCT TGTGTTGAAT AGTATCTTAT   9409

ATATTTTAAT ATGTCTTATT AGTCTCATTA CTTAGGCGAA CGACAAAGTG AGGTCACCTC   9469

GGTCTAATTC TCCTATGTAG TGCGAG                                        9495
```

(3) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 792

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Double

(D) TOPOLOGY: Circular

(ii) MOLECULE TYPE: cDNA to genomic RNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(v) FRAGMENT TYPE: N/A

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Tobacco Etch Virus

(B) STRAIN: Highly Aphid Transmitted

(C) INDIVIDUAL ISOLATE: N/A

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: No

(B) CLONE: pTC:FL

(viii) POSITION IN GENOME: N/A

(ix) FEATURE:

(A) NAME/KEY: Mutations (AGT-ATG) introduced into nucleotides corresponding to genomic nucleotides 8518-8520 of SEQ ID No. 1, to create initiating methionine codon.

(B) LOCATION: Nucleotides 1-3 of SEQ ID No. 2

(C) IDENTIFICATION METHOD: --

(D) OTHER INFORMATION: SEQ ID NO: 2 is the modified Tobacco Etch Virus coat protein gene present in pTC:FL.

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Allison et al.

(B) TITLE: The nucleotide sequence of the coding region of Tobacco Etch Virus Genomic RNA: Evidence for the Synthesis of a Single Polyprotein

(C) JOURNAL: Virology

(D) VOLUME: 154

(E) ISSUE: --

(F) PAGES: 9-20

(A) AUTHORS: Lindbo and Dougherty

(B) TITLE: Untranslatable Transcripts of the tobacco etch virus coat protein gene sequence can interfere with tobacco etch virus replication in Transgenic Plants and Protoplasts

(C) JOURNAL: Virology

(D) VOLUME: 189

(E) ISSUE: --

(F) PAGES: 725-733

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
                                                                ATG GGC ACT        9
                                                                Met Gly Thr
                                                                 1

GTG GAT GCT GGT GCT GAC GCT GGT AAG AAG AAA GAT CAA AAG GAT GAT       57
Val Asp Ala Gly Ala Asp Ala Gly Lys Lys Lys Asp Gln Lys Asp Asp
     5                  10                  15

AAA GTC GCT GAG CAG GCT TCA AAG GAT AGG GAT GTT AAT GCT GGA ACT      105
Lys Val Ala Glu Gln Ala Ser Lys Asp Arg Asp Val Asn Ala Gly Thr
 20                  25                  30                  35

TCA GGA ACA TTC TCA GTT CCA CGA ATA AAT GCT ATG GCC ACA AAA CTT      153
Ser Gly Thr Phe Ser Val Pro Arg Ile Asn Ala Met Ala Thr Lys Leu
                 40                  45                  50

CAA TAT CCA AGG ATG AGG GGA GAG GTG GTT GTA AAC TTG AAT CAC CTT      201
Gln Tyr Pro Arg Met Arg Gly Glu Val Val Val Asn Leu Asn His Leu
             55                  60                  65

TTA GGA TAC AAG CCA CAG CAA ATT GAT TTG TCA AAT GCT CGA GCC ACA      249
Leu Gly Tyr Lys Pro Gln Gln Ile Asp Leu Ser Asn Ala Arg Ala Thr
         70                  75                  80

CAT GAG CAG TTT GCC GCG TGG CAT CAG GCA GTG ATG ACA GCC TAT GGA      297
His Glu Gln Phe Ala Ala Trp His Gln Ala Val Met Thr Ala Tyr Gly
 85                  90                  95

GTG AAT GAA GAG CAA ATG AAA ATA TTG CTA AAT GGA TTT ATG GTG TGG      345
Val Asn Glu Glu Gln Met Lys Ile Leu Leu Asn Gly Phe Met Val Trp
100                 105                 110                 115

TGC ATA GAA AAT GGG ACT TCC CCA AAT TTG AAC GGA ACT TGG GTT ATG      393
Cys Ile Glu Asn Gly Thr Ser Pro Asn Leu Asn Gly Thr Trp Val Met
                120                 125                 130
```

```
ATG GAT GGT GAG GAT CAA GTT TCA TAC CCG CTG AAA CCA ATG GTT GAA      441
Met Asp Gly Glu Asp Gln Val Ser Tyr Pro Leu Lys Pro Met Val Glu
            135                 140                 145

AAC GCG CAG CCA ACA CTG AGG CAA ATT ATG ACA CAC TTC AGT GAC CTG      489
Asn Ala Gln Pro Thr Leu Arg Gln Ile Met Thr His Phe Ser Asp Leu
        150                 155                 160

GCT GAA GCG TAT ATT GAG ATG AGG AAT AGG GAG CGA CCA TAC ATG CCT      537
Ala Glu Ala Tyr Ile Glu Met Arg Asn Arg Glu Arg Pro Tyr Met Pro
    165                 170                 175

AGG TAT GGT CTA CAG AGA AAC ATT ACA GAC ATG AGT TTG TCA CGC TAT      585
Arg Tyr Gly Leu Gln Arg Asn Ile Thr Asp Met Ser Leu Ser Arg Tyr
180                 185                 190                 195

GCG TTC GAC TTC TAT GAG CTA ACT TCA AAA ACA CCT GTT AGA GCG AGG      633
Ala Phe Asp Phe Tyr Glu Leu Thr Ser Lys Thr Pro Val Arg Ala Arg
                200                 205                 210

GAG GCG CAT ATG CAA ATG AAA GCT GCT GCA GTA CGA AAC AGT GGA ACT      681
Glu Ala His Met Gln Met Lys Ala Ala Ala Val Arg Asn Ser Gly Thr
            215                 220                 225

AGG TTA TTT GGT CTT GAT GGC AAC GTG GGT ACT GCA GAG GAA GAC ACT      729
Arg Leu Phe Gly Leu Asp Gly Asn Val Gly Thr Ala Glu Glu Asp Thr
        230                 235                 240

GAA CGG CAC ACA GCG CAC GAT GTG AAC CGT AAC ATG CAC ACA CTA TTA      777
Glu Arg His Thr Ala His Asp Val Asn Arg Asn Met His Thr Leu Leu
    245                 250                 255

GGG GTC CGC CAG TGA                                                 792
Gly Val Arg Gln
260
```

(4) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 793
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Double
(D) TOPOLOGY: Circular

(ii) MOLECULE TYPE: cDNA to genomic RNA
(iii) HYPOTHETICAL: No
(iv) ANTI-SENSE: No
(v) FRAGMENT TYPE: N/A
(vi) ORIGINAL SOURCE:

(A) ORGANISM: Tobacco Etch Virus
(B) STRAIN: Highly Aphid Transmitted
(C) INDIVIDUAL ISOLATE: N/A

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: No
(B) CLONE: pTC:RC

(viii) POSITION IN GENOME: N/A
(ix) FEATURE:

(A) NAME/KEY: Mutation of AGT-GGC (Ser-Gly) to ATG-GCC (Met-Ser)

(B) LOCATION: Nucleotides 1-6 of SEQ ID NO. 3 (corresponding to nucleotides 8518-8523 of SEQ ID NO. 1)

(A) NAME/KEY: Frameshift mutation (insertion of T) producing stop codon

(B) LOCATION: Nucleotide 13 of SEQ ID No. 3 (corresponding to position between nucleotides 8529 and 8530 of SEQ. ID No. 1)

(D) OTHER INFORMATION: SEQ ID No: 3 is the modified Tobacco Etch Virus coat protein gene present in pTC:RC.

(x) PUBLICATION INFORMATION:

(A) AUTHORS: J. A. Lindbo and W. G. Dougherty

(B) TITLE: Pathogen-Derived Resistance to a Potyvirus: Immune and Resistant Phenotypes in Transgenic Tobacco Expressing Altered Forms of a Potyvirus Coat Protein Nucleotide Sequence

(C) JOURNAL: Molecular Plant-Microbe Interactions

(D) VOLUME: 5

(E) ISSUE: 2

(F) PAGES: 144-153

(A) AUTHORS: J. A. Lindbo and W. G. Dougherty

(B) TITLE: Untranslatable Transcripts of the Tobacco Etch Virus Coat Protein Gene Sequence Can Interfere with Tobacco Etch Virus Replication in Transgenic Plants and Protoplasts

(C) JOURNAL: Virology

(D) VOLUME: 189

(E) ISSUE: --

(F) PAGES: 725-733

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
                                                ATG GCC ACT          9
                                                Met Ser Thr

GTG TGA TGA TGGTGCTAGC GCTGGTAAGA AGAAAGATCA AAAGGATGAT          58
Val
```

```
AAAGTCGCTG AGCAGGCTTC AAAGGATAGG GATGTTAATG CTGGAACTTC      108

AGGAACATTC TCAGTTCCAC GAATAAATGC TATGGCCACA AAACTTCAAT      158

ATCCAAGGAT GAGGGGAGAG GTGGTTGTAA ACTTGAATCA CCTTTTAGGA      208

TACAAGCCAC AGCAAATTGA TTTGTCAAAT GCTCGAGCCA CACATGAGCA      258

GTTTGCCGCG TGGCATCAGG CAGTGATGAC AGCCTATGGA GTGAATGAAG      308

AGCAAATGAA AATATTGCTA AATGGATTTA TGGTGTGGTG CATAGAAAAT      358

GGGACTTCCC CAAATTTGAA CGGAACTTGG GTTATGATGG ATGGTGAGGA      408

TCAAGTTTCA TACCCGCTGA AACCAATGGT TGAAAACGCG CAGCCAACAC      458

TGAGGCAAAT TATGACACAC TTCAGTGACC TGGCTGAAGC GTATATTGAG      508

ATGAGGAATA GGGAGCGACC ATACATGCCT AGGTATGGTC TACAGAGAAA      558

CATTACAGAC ATGAGTTTGT CACGCTATGC GTTCGACTTC TATGAGCTAA      608

CTTCAAAAAC ACCTGTTAGA GCGAGGGAGG CGCATATGCA AATGAAAGCT      658

GCTGCAGTAC GAAACAGTGG AACTAGGTTA TTTGGTCTTG ATGGCAACGT      708

GGGTACTGCA GAGGAAGACA CTGAACGGCA CACAGCGCAC GATGTGAACC      758

GTAACATGCA CACACTATTA GGGGTCCGCC AGTGA                      793
```

(5) INFORMATION FOR SEQ ID NO: 4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 792
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: Double
(D) TOPOLOGY: Circular

(ii) MOLECULE TYPE: cDNA to genomic RNA
(iii) HYPOTHETICAL: No
(iv) ANTI-SENSE: Yes
(v) FRAGMENT TYPE: N/A
(vi) ORIGINAL SOURCE:

(A) ORGANISM: Tobacco Etch Virus
(B) STRAIN: Highly Aphid Transmitted
(C) INDIVIDUAL ISOLATE: N/A

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: No
(B) CLONE: pTC:AS

(viii) POSITION IN GENOME: N/A
(ix) FEATURE:

(A) NAME/KEY: --
(B) LOCATION: --
(C) IDENTIFICATION METHOD: --

(D) OTHER INFORMATION: SEQ ID No. 4 is the modified Tobacco Etch Virus Coat protein gene present in pTC:AS. It is the inverse complement of SEQ ID No. 2.

(x) PUBLICATION INFORMATION:

(A) AUTHORS: J. A. Lindbo and W. G. Dougherty

(B) TITLE: Untranslatable Transcripts of the Tobacco Etch Virus Coat Protein Gene Sequence Can Interfere with Tobacco Etch Virus Replication in Transgenic Plants and Protoplasts

(C) JOURNAL: Virology

(D) VOLUME: 189

(E) ISSUE: --

(F) PAGES: 725-733

(A) AUTHORS: J. A. Lindbo and W. G. Dougherty

(B) TITLE: Pathogen-Derived Resistance to a Potyvirus: Immune and Resistant Phenotypes in Transgenic Tobacco Expressing Altered Forms of a Potyvirus Coat Protein Nucleotide Sequence

(C) JOURNAL: Molecular Plant-Microbe Interactions

(D) VOLUME: 5

(E) ISSUE: 2

(F) PAGES: 144-153

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
TCACTGGCGG ACCCCTAATA GTGTGTGCAT GTTACGGTTC ACATCGTGCG CTGTGTGCCG      60
TTCAGTGTCT TCCTCTGCAG TACCCACGTT GCCATCAAGA CCAAATAACC TAGTTCCACT     120
GTTTCGTACT GCAGCAGCTT TCATTTGCAT ATGCGCCTCC CTCGCTGTAA CAGGTGTTTT     180
TGAAGTTAGC TCATAGAAGT CGAACGCATA GCGTGACAAA CTCATGTCTG TAATGTTTCT     240
CTGTAGACCA TACCTAGGCA TGTATGGTCG CTCCCTATTC CTCATCTCAA TATACGCTTC     300
AGCCAGGTCA CTGAAGTGTG TCATAATTTG CCTCAGTGTT GGCTGCGCGT TTTCAACCAT     360
TGGTTTCAGC GGGTATGAAA CTTGATCCTC ACCATCCATC ATAACCCAAG TTCCGTTCAA     420
ATTTGGGGAA GTCCCATTTT CTATGCACCA CACCATAAAT CCATTTAGCA ATATTTTCAT     480
TTGCTCTTCA TTCACTCCAT AGGCTGTCAT CACTGCCTGA TGCCACGCGG CAAACTGCTC     540
ATGTGTGGCT CGAGCATTTG ACAAATCAAT TTGCTGTGGC TTGTATCCTA AAAGGTGATT     600
CAAGTTTACA ACCACCTCTC CCCTCATCCT TGGATATTGA AGTTTTGTGG CCATAGCATT     660
TATTCGTGGA ACTGAGAATG TTCCTGAAGT TCCAGCATTA ACATCCCTAT CCTTTGAAGC     720
CTGCTCAGCG ACTTTATCAT CCTTTTGATC TTTCTTCTTA CCAGCGTCAG CACCAGCATC     780
CACAGTGCCC AT                                                        792
```

Claims

1. A method of producing a plant with a reduced susceptibility to viral infection, comprising:

   transforming plant cells with a DNA molecule that encodes untranslatable plus-sense viral RNA molecule wherein the untranslatable plus-sense viral RNA molecule is derived from the nucleotide sequence of a plant virus gene; and

   regenerating a plant comprising the transformed plant cell.

2. A transgenic plant produced according to the method of Claim 1.

3. The method of Claim 1 wherein the untranslatable plus-sense viral RNA molecule is derived from a viral coat protein gene.

4. The method of Claim 1 wherein the untranslatable plus-sense viral RNA molecule is derived from a potyvirus.

5. A DNA molecule useful for producing virus resistant plants comprising a promoter operably linked to a DNA molecule encoding an untranslatable plus-sense viral RNA molecule, derived from the nucleotide sequence of a plant virus gene.

6. The method of any one of Claims 1, 3 or 4 wherein the untranslatable plus-sense viral RNA molecule contains at least one mutation that renders the RNA molecule untranslatable, and expression of the untranslatable plus-sense viral RNA molecule within the plant reduces the susceptibility of the plant to virus infection;
   and wherein the method further comprises the step of selecting a plant that shows a reduced susceptibility to infection by the virus.

Patentansprüche

1. Ein Verfahren zur Herstellung einer Pflanze mit einer reduzierten Anfälligkeit für eine virale Infektion, umfassend:

   Transformation von Pflanzenzellen mit einem DNA-Molekül, das für ein nicht-translatierbares plus-strang virales RNA-Molekül kodiert, **dadurch gekennzeichnet, daß** das nicht-translatierbare plus-strang virale RNA-Molekül von der Nukleotidsequenz eines Pflanzenvirusgens abgeleitet ist; und

   Regeneration einer Pflanze, beinhaltend die transformierte Pflanzenzelle.

2. Eine transgene Pflanze, hergestellt entsprechend des Verfahrens nach Anspruch 1.

3. Das Verfahren nach Anspruch 1, **dadurch gekennzeichnet, daß** das nicht-translatierbare plus-strang virale RNA-Molekül von einem viralen Hüllproteingen abgeleitet ist.

4. Das Verfahren nach Anspruch 1, **dadurch gekennzeichnet, daß** das nicht-translatierbare plus-strang virale RNA-Molekül von einem Potyvirus abgeleitet ist.

5. Ein DNA-Molekül, verwendbar zur Herstellung virusresistenter Planzen, umfassend einen Promoter, wirksam verbunden mit einem DNA-Molekül, das für ein nicht-translatierbares plus-strang virales RNA-Molekül kodiert, welches von der Nukleotidsequenz eines Pflanzenvirusgens abgeleitet ist.

6. Das Verfahren nach einem der Ansprüche 1, 3 oder 4, **dadurch gekennzeichnet, daß** das nicht-translatierbare plus-strang virale RNA-Molekül mindestens eine Mutation enthält, die das RNA-Molekül nicht-translatierbar macht, und daß die Expression des nicht-translatierbaren plus-strang viralen RNA-Moleküls in der Pflanze zu einer reduzierten Anfälligkeit der Pflanze gegen Virusinfektion führt;
   und, **dadurch gekennzeichnet, daß** das Verfahren weiterhin den Schritt der Selektion einer Pflanze umfaßt, die eine reduzierte Anfälligkeit für eine Infektion durch das Virus zeigt.

Revendications

1. Méthode de production d'une plante avec une sensibilité réduite à l'infection virale, comprenant les étapes consistant :

   à transformer des cellules végétales avec une molécule d'ADN qui code une molécule d'ARN viral sens plus non traduisible, la molécule d'ARN viral sens plus non traduisible étant dérivée de la séquence nucléotidique d'un gène de virus de plante ; et
   à régénérer une plante comprenant la cellule végétale transformée.

2. Plante transgénique produite selon la méthode de la revendication 1.

3. Méthode selon la revendication 1, dans laquelle la molécule d'ARN viral sens plus non traduisible est dérivée d'un gène de protéine de capside virale.

4. Méthode selon la revendication 1, dans laquelle la molécule d'ARN viral sens plus non traduisible est dérivée d'un potyvirus.

5. Molécule d'ADN utile pour produire des plantes résistantes à des virus, comprenant un promoteur lié de façon opérationnelle à une molécule d'ADN codant une molécule d'ARN viral sens plus non traduisible, dérivée de la séquence nucléotidique d'un gène de virus de plante.

6. Méthode selon l'une quelconque des revendications 1, 3 ou 4, dans laquelle la molécule d'ARN viral sens plus non traduisible contient au moins une mutation qui rend la molécule d'ARN non traduisible, et l'expression de la molécule d'ARN viral sens plus non traduisible dans la plante réduit la sensibilité de la plante à une infection virale ; et la méthode comprenant en outre l'étape consistant à sélectionner une plante qui montre une sensibilité réduite à une infection par le virus.

```
NAAATAACAA ATCTCAACAC AACATATACA AAACAAACGA ATCTCAAGCA ATCAAGCATT      60

CTACTTCTAT TGCAGCAATT TAAATCATTT CTTTTAAAGC AAAAGCAATT TTCTGAAAAT     120

TTTCACCATT TACGAACGAT AGCA ATG GCA CTG ATC TTT GGC ACA GTC AAC GCT     174
                           Met Ala Leu Ile Phe Gly Thr Val Asn Ala
                             1               5                  10

AAC ATC CTG AAG GAA GTG TTC GGT GGA GCT CGT ATG GCT TGC GTT ACC       222
Asn Ile Leu Lys Glu Val Phe Gly Gly Ala Arg Met Ala Cys Val Thr
                  15                  20                  25

AGC GCA CAT ATG GCT GGA GCG AAT GGA AGC ATT TTG AAG AAG GCA GAA       270
Ser Ala His Met Ala Gly Ala Asn Gly Ser Ile Leu Lys Lys Ala Glu
             30                  35                  40

GAG ACC TCT CGT GCA ATC ATG CAC AAA CCA GTG ATC TTC GGA GAA GAC       318
Glu Thr Ser Arg Ala Ile Met His Lys Pro Val Ile Phe Gly Glu Asp
         45                  50                  55

TAC ATT ACC GAG GCA GAC TTG CCT TAC ACA CCA CTC CAT TTA GAG GTC       366
Tyr Ile Thr Glu Ala Asp Leu Pro Tyr Thr Pro Leu His Leu Glu Val
     60                  65                  70

GAT GCT GAA ATG GAG CGG ATG TAT TAT CTT GGT CGT CGC GCG CTC ACC       414
Asp Ala Glu Met Glu Arg Met Tyr Tyr Leu Gly Arg Arg Ala Leu Thr
 75                  80                  85                  90

CAT GGC AAG AGA CGC AAA GTT TCT GTG AAT AAC AAG AGG AAC AGG AGA       462
His Gly Lys Arg Arg Lys Val Ser Val Asn Asn Lys Arg Asn Arg Arg
                 95                 100                 105

AGG AAA GTG GCC AAA ACG TAC GTG GGG CGT GAT TCC ATT GTT GAG AAG       510
Arg Lys Val Ala Lys Thr Tyr Val Gly Arg Asp Ser Ile Val Glu Lys
            110                 115                 120

ATT GTA GTG CCC CAC ACC GAG AGA AAG GTT GAT ACC ACA GCA GCA GTG       558
Ile Val Val Pro His Thr Glu Arg Lys Val Asp Thr Thr Ala Ala Val
        125                 130                 135

GAA GAC ATT TGC AAT GAA GCT ACC ACT CAA CTT GTG CAT AAT AGT ATG       606
Glu Asp Ile Cys Asn Glu Ala Thr Thr Gln Leu Val His Asn Ser Met
    140                 145                 150

CCA AAG CGT AAG AAG CAG AAA AAC TTC TTG CCC GCC ACT TCA CTA AGT       654
Pro Lys Arg Lys Lys Gln Lys Asn Phe Leu Pro Ala Thr Ser Leu Ser
155                 160                 165                 170

AAC GTG TAT GCC CAA ACT TGG AGC ATA GTG CGC AAA CGC CAT ATG CAG       702
Asn Val Tyr Ala Gln Thr Trp Ser Ile Val Arg Lys Arg His Met Gln
                175                 180                 185

GTG GAG ATC ATT AGC AAG AAG AGC GTC CGA GCG AGG GTC AAG AGA TTT       750
Val Glu Ile Ile Ser Lys Lys Ser Val Arg Ala Arg Val Lys Arg Phe
            190                 195                 200

GAG GGC TCG GTG CAA TTG TTC GCA AGT GTG CGT CAC ATG TAT GGC GAG       798
Glu Gly Ser Val Gln Leu Phe Ala Ser Val Arg His Met Tyr Gly Glu
        205                 210                 215

AGG AAA AGG GTG GAC TTA CGT ATT GAC AAC TGG CAG CAA GAG ACA CTT       846
Arg Lys Arg Val Asp Leu Arg Ile Asp Asn Trp Gln Gln Glu Thr Leu
    220                 225                 230
```

FIG. 1

```
CTA GAC CTT GCT AAA AGA TTT AAG AAT GAG AGA GTG GAT CAA TCG AAG      894
Leu Asp Leu Ala Lys Arg Phe Lys Asn Glu Arg Val Asp Gln Ser Lys
235                 240                 245                 250

CTC ACT TTT GGT TCA AGT GGC CTA GTT TTG AGG CAA GGC TCG TAC GGA      942
Leu Thr Phe Gly Ser Ser Gly Leu Val Leu Arg Gln Gly Ser Tyr Gly
                255                 260                 265

CCT GCG CAT TGG TAT CGA CAT GGT ATG TTC ATT GTA CGC GGT CGG TCG      990
Pro Ala His Trp Tyr Arg His Gly Met Phe Ile Val Arg Gly Arg Ser
            270                 275                 280

GAT GGG ATG TTG GTG GAT GCT CGT GCG AAG GTA ACG TTC GCT GTT TGT     1038
Asp Gly Met Leu Val Asp Ala Arg Ala Lys Val Thr Phe Ala Val Cys
        285                 290                 295

CAC TCA ATG ACA CAT TAT AGC GAC AAA TCA ATC TCT GAG GCA TTC TTC     1086
His Ser Met Thr His Tyr Ser Asp Lys Ser Ile Ser Glu Ala Phe Phe
    300                 305                 310

ATA CCA TAC TCT AAG AAA TTC TTG GAG TTG AGA CCA GAT GGA ATC TCC     1134
Ile Pro Tyr Ser Lys Lys Phe Leu Glu Leu Arg Pro Asp Gly Ile Ser
315                 320                 325                 330

CAT GAG TGT ACA AGA GGA GTA TCA GTT GAG CGG TGC GGT GAG GTG GCT     1182
His Glu Cys Thr Arg Gly Val Ser Val Glu Arg Cys Gly Glu Val Ala
                335                 340                 345

GCA ATC CTG ACA CAA GCA CTT TCA CCG TGT GGT AAG ATC ACA TGC AAA     1230
Ala Ile Leu Thr Gln Ala Leu Ser Pro Cys Gly Lys Ile Thr Cys Lys
            350                 355                 360

CGT TGC ATG GTT GAA ACA CCT GAC ATT GTT GAG GGT GAG TCG GGA GAA     1278
Arg Cys Met Val Glu Thr Pro Asp Ile Val Glu Gly Glu Ser Gly Glu
        365                 370                 375

AGT GTC ACC AAC CAA GGT AAG CTC CTA GCA ATG CTG AAA GAA CAG TAT     1326
Ser Val Thr Asn Gln Gly Lys Leu Leu Ala Met Leu Lys Glu Gln Tyr
    380                 385                 390

CCA GAT TTC CCA ATG GCC GAG AAA CTA CTC ACA AGG TTT TTG CAA CAG     1374
Pro Asp Phe Pro Met Ala Glu Lys Leu Leu Thr Arg Phe Leu Gln Gln
395                 400                 405                 410

AAA TCA CTA GTA AAT ACA AAT TTG ACA GCC TGC GTG AGC GTC AAA CAA     1422
Lys Ser Leu Val Asn Thr Asn Leu Thr Ala Cys Val Ser Val Lys Gln
                415                 420                 425

CTC ATT GGT GAC CGC AAA CAA GCT CCA TTC ACA CAC GTA CTG GCT GTC     1470
Leu Ile Gly Asp Arg Lys Gln Ala Pro Phe Thr His Val Leu Ala Val
            430                 435                 440

AGC GAA ATT CTG TTT AAA GGC AAT AAA CTA ACA GGG GCT GAT CTC GAA     1518
Ser Glu Ile Leu Phe Lys Gly Asn Lys Leu Thr Gly Ala Asp Leu Glu
        445                 450                 455

GAG GCA AGC ACA CAT ATG CTT GAA ATA GCA AGG TTC TTG AAC AAT CGC     1566
Glu Ala Ser Thr His Met Leu Glu Ile Ala Arg Phe Leu Asn Asn Arg
    460                 465                 470

ACT GAA AAT ATG CGC ATT GGC CAC CTT GGT TCT TTC AGA AAT AAA ATC     1614
Thr Glu Asn Met Arg Ile Gly His Leu Gly Ser Phe Arg Asn Lys Ile
475                 480                 485                 490
```

FIG. 1

```
TCA TCG AAG GCC CAT GTG AAT AAC GCA CTC ATG TGT GAT AAT CAA CTT      1662
Ser Ser Lys Ala His Val Asn Asn Ala Leu Met Cys Asp Asn Gln Leu
                 495                 500                 505

GAT CAG AAT GGG AAT TTT ATT TGG GGA CTA AGG GGT GCA CAC GCA AAG      1710
Asp Gln Asn Gly Asn Phe Ile Trp Gly Leu Arg Gly Ala His Ala Lys
            510                 515                 520

AGG TTT CTT AAA GGA TTT TTC ACT GAG ATT GAC CCA AAT GAA GGA TAC      1758
Arg Phe Leu Lys Gly Phe Phe Thr Glu Ile Asp Pro Asn Glu Gly Tyr
        525                 530                 535

GAT AAG TAT GTT ATC AGG AAA CAT ATC AGG GGT AGC AGA AAG CTA GCA      1806
Asp Lys Tyr Val Ile Arg Lys His Ile Arg Gly Ser Arg Lys Leu Ala
    540                 545                 550

ATT GGC AAT TTG ATA ATG TCA ACT GAC TTC CAG ACG CTC AGG CAA CAA      1854
Ile Gly Asn Leu Ile Met Ser Thr Asp Phe Gln Thr Leu Arg Gln Gln
555                 560                 565                 570

ATT CAA GGC GAA ACT ATT GAG CGT AAA GAA ATT GGG AAT CAC TGC ATT      1902
Ile Gln Gly Glu Thr Ile Glu Arg Lys Glu Ile Gly Asn His Cys Ile
                575                 580                 585

TCA ATG CGG AAT GGT AAT TAC GTG TAC CCA TGT TGT TGT GTT ACT CTT      1950
Ser Met Arg Asn Gly Asn Tyr Val Tyr Pro Cys Cys Cys Val Thr Leu
            590                 595                 600

GAA GAT GGT AAG GCT CAA TAT TCG GAT CTA AAG CAC CCA ACG AAG AGA      1998
Glu Asp Gly Lys Ala Gln Tyr Ser Asp Leu Lys His Pro Thr Lys Arg
        605                 610                 615

CAT CTG GTC ATT GGC AAC TCT GGC GAT TCA AAG TAC CTA GAC CTT CCA      2046
His Leu Val Ile Gly Asn Ser Gly Asp Ser Lys Tyr Leu Asp Leu Pro
    620                 625                 630

GTT CTC AAT GAA GAG AAA ATG TAT ATA GCT AAT GAA GGT TAT TGC TAC      2094
Val Leu Asn Glu Glu Lys Met Tyr Ile Ala Asn Glu Gly Tyr Cys Tyr
635                 640                 645                 650

ATG AAC ATT TTC TTT GCT CTA CTA GTG AAT GTC AAG GAA GAG GAT GCA      2142
Met Asn Ile Phe Phe Ala Leu Leu Val Asn Val Lys Glu Glu Asp Ala
                655                 660                 665

AAG GAC TTC ACC AAG TTT ATA AGG GAC ACA ATT GTT CCA AAG CTT GGA      2190
Lys Asp Phe Thr Lys Phe Ile Arg Asp Thr Ile Val Pro Lys Leu Gly
            670                 675                 680

GCG TGG CCA ACA ATG CAA GAT GTT GCA ACT GCA TGC TAC TTA CTT TCC      2238
Ala Trp Pro Thr Met Gln Asp Val Ala Thr Ala Cys Tyr Leu Leu Ser
        685                 690                 695

ATT CTT TAC CCA GAT GTC CTG AGA GCT GAA CTA CCC AGA ATT TTG GTT      2286
Ile Leu Tyr Pro Asp Val Leu Arg Ala Glu Leu Pro Arg Ile Leu Val
    700                 705                 710

GAT CAT GAC AAC AAA ACA ATG CAT GTT TTG GAT TCG TAT GGG TCT AGA      2334
Asp His Asp Asn Lys Thr Met His Val Leu Asp Ser Tyr Gly Ser Arg
715                 720                 725                 730

ACG ACA GGA TAC CAC ATG TTG AAA ATG AAC ACA ACA TCC CAG CTA ATT      2382
Thr Thr Gly Tyr His Met Leu Lys Met Asn Thr Thr Ser Gln Leu Ile
                735                 740                 745
```

FIG. 1

```
GAA TTC GTT CAT TCA GGT TTG GAA TCC GAA ATG AAA ACT TAC AAT GTT      2430
Glu Phe Val His Ser Gly Leu Glu Ser Glu Met Lys Thr Tyr Asn Val
            750                 755                 760

GGA GGG ATG AAC CGA GAT GTG GTC ACA CAA GGT GCA ATT GAG ATG TTG      2478
Gly Gly Met Asn Arg Asp Val Val Thr Gln Gly Ala Ile Glu Met Leu
        765                 770                 775

ATC AAG TCT ATA TAC AAA CCA CAT CTC ATG AAG CAG TTA CTT GAG GAA      2526
Ile Lys Ser Ile Tyr Lys Pro His Leu Met Lys Gln Leu Leu Glu Glu
    780                 785                 790

GAG CCA TAC ATA ATT GTC CTG GCA ATA GTC TCC CCT TCA ATT TTA ATT      2574
Glu Pro Tyr Ile Ile Val Leu Ala Ile Val Ser Pro Ser Ile Leu Ile
795                 800                 805                 810

GCC ATG TAC AAC TCT GGA ACT TTT GAG CAG GCG TTA CAA ATG TGG TTG      2622
Ala Met Tyr Asn Ser Gly Thr Phe Glu Gln Ala Leu Gln Met Trp Leu
                815                 820                 825

CCA AAT ACA ATG AGG TTA GCT AAC CTC GCT GCC ATC TTG TCA GCC TTA      2670
Pro Asn Thr Met Arg Leu Ala Asn Leu Ala Ala Ile Leu Ser Ala Leu
            830                 835                 840

GCG CAA AAG TTA ACT TTG GCA GAT TTG TTC GTC CAG CAG CGT AAT TTG      2718
Ala Gln Lys Leu Thr Leu Ala Asp Leu Phe Val Gln Gln Arg Asn Leu
        845                 850                 855

ATT AAT GAG TAT GCG CAG GTA ATT TTG GAC AAT CTG ATT GAC GGT GTC      2766
Ile Asn Glu Tyr Ala Gln Val Ile Leu Asp Asn Leu Ile Asp Gly Val
    860                 865                 870

AGG GTT AAT CAT TCG CTA TCC CTA GCA ATG GAA ATT GTT ACT ATT AAG      2814
Arg Val Asn His Ser Leu Ser Leu Ala Met Glu Ile Val Thr Ile Lys
875                 880                 885                 890

CTG GCC ACC CAA GAG ATG GAC ATG GCG TTG AGG GAA GGT GGC TAT GCT      2862
Leu Ala Thr Gln Glu Met Asp Met Ala Leu Arg Glu Gly Gly Tyr Ala
                895                 900                 905

GTG ACC TCT GAA AAG GTG CAT GAA ATG TTG GAA AAA AAC TAT GTA AAG      2910
Val Thr Ser Glu Lys Val His Glu Met Leu Glu Lys Asn Tyr Val Lys
            910                 915                 920

GCT TTG AAG GAT GCA TGG GAC GAA TTA ACT TGG TTG GAA AAA TTC TCC      2958
Ala Leu Lys Asp Ala Trp Asp Glu Leu Thr Trp Leu Glu Lys Phe Ser
        925                 930                 935

GCA ATC AGG CAT TCA AGA AAG CTC TTG AAA TTT GGG CGA AAG CCT TTA      3006
Ala Ile Arg His Ser Arg Lys Leu Leu Lys Phe Gly Arg Lys Pro Leu
    940                 945                 950

ATC ATG AAA AAC ACC GTA GAT TGC GGC GGA CAT ATA GAC TTG TCT GTG      3054
Ile Met Lys Asn Thr Val Asp Cys Gly Gly His Ile Asp Leu Ser Val
955                 960                 965                 970

AAA TCG CTT TTC AAG TTC CAC TTG GAA CTC CTG AAG GGA ACC ATC TCA      3102
Lys Ser Leu Phe Lys Phe His Leu Glu Leu Leu Lys Gly Thr Ile Ser
                975                 980                 985

AGA GCC GTA AAT GGT GGC GCA AGA AAG GTA AGA GTA GCG AAG AAT GCC      3150
Arg Ala Val Asn Gly Gly Ala Arg Lys Val Arg Val Ala Lys Asn Ala
            990                 995                1000
```

FIG. 1

```
ATG ACA AAA GGG GTT TTT CTC AAA ATC TAC AGC ATG CTT CCT GAC GTC     3198
Met Thr Lys Gly Val Phe Leu Lys Ile Tyr Ser Met Leu Pro Asp Val
        1005                1010                1015

TAC AAG TTT ATC ACA GTC TCG AGT GTC CTT TCC TTG TTG TTG ACA TTC     3246
Tyr Lys Phe Ile Thr Val Ser Ser Val Leu Ser Leu Leu Leu Thr Phe
    1020                1025                1030

TTA TTT CAA ATT GAC TGC ATG ATA AGG GCA CAC CGA GAG GCG AAG GTT     3294
Leu Phe Gln Ile Asp Cys Met Ile Arg Ala His Arg Glu Ala Lys Val
1035                1040                1045                1050

GCT GCA CAG TTG CAG AAA GAG AGC GAG TGG GAC AAT ATC ATC AAT AGA     3342
Ala Ala Gln Leu Gln Lys Glu Ser Glu Trp Asp Asn Ile Ile Asn Arg
                1055                1060                1065

ACT TTC CAG TAT TCT AAG CTT GAA AAT CCT ATT GGC TAT CGC TCT ACA     3390
Thr Phe Gln Tyr Ser Lys Leu Glu Asn Pro Ile Gly Tyr Arg Ser Thr
            1070                1075                1080

GCG GAG GAA AGA CTC CAA TCA GAA CAC CCC GAG GCT TTC GAG TAC TAC     3438
Ala Glu Glu Arg Leu Gln Ser Glu His Pro Glu Ala Phe Glu Tyr Tyr
        1085                1090                1095

AAG TTT TGC ATT GGA AAG GAA GAC CTC GTT GAA CAG GCA AAA CAA CCG     3486
Lys Phe Cys Ile Gly Lys Glu Asp Leu Val Glu Gln Ala Lys Gln Pro
    1100                1105                1110

GAG ATA GCA TAC TTT GAA AAG ATT ATA GCT TTC ATC ACA CTT GTA TTA     3534
Glu Ile Ala Tyr Phe Glu Lys Ile Ile Ala Phe Ile Thr Leu Val Leu
1115                1120                1125                1130

ATG GCT TTT GAC GCT GAG CGG AGT GAT GGA GTG TTC AAG ATA CTC AAT     3582
Met Ala Phe Asp Ala Glu Arg Ser Asp Gly Val Phe Lys Ile Leu Asn
                1135                1140                1145

AAG TTC AAA GGA ATA CTG AGC TCA ACG GAG AGG GAG ATC ATC TAC ACG     3630
Lys Phe Lys Gly Ile Leu Ser Ser Thr Glu Arg Glu Ile Ile Tyr Thr
            1150                1155                1160

CAG AGT TTG GAT GAT TAC GTT ACA ACC TTT GAT GAC AAT ATG ACA ATC     3678
Gln Ser Leu Asp Asp Tyr Val Thr Thr Phe Asp Asp Asn Met Thr Ile
        1165                1170                1175

AAC CTC GAG TTG AAT ATG GAT GAA CTC CAC AAG ACG AGC CTT CCT GGA     3726
Asn Leu Glu Leu Asn Met Asp Glu Leu His Lys Thr Ser Leu Pro Gly
    1180                1185                1190

GTC ACT TTT AAG CAA TGG TGG AAC AAC CAA ATC AGC CGA GGC AAC GTG     3774
Val Thr Phe Lys Gln Trp Trp Asn Asn Gln Ile Ser Arg Gly Asn Val
1195                1200                1205                1210

AAG CCA CAT TAT AGA ACT GAG GGG CAC TTC ATG GAG TTT ACC AGA GAT     3822
Lys Pro His Tyr Arg Thr Glu Gly His Phe Met Glu Phe Thr Arg Asp
                1215                1220                1225

ACT GCG GCA TCG GTT GCC AGC GAG ATA TCA CAC TCA CCC GCA AGA GAT     3870
Thr Ala Ala Ser Val Ala Ser Glu Ile Ser His Ser Pro Ala Arg Asp
            1230                1235                1240

TTT CTT GTG AGA GGT GCT GTT GGA TCT GGA AAA TCC ACA GGA CTT CCA     3918
Phe Leu Val Arg Gly Ala Val Gly Ser Gly Lys Ser Thr Gly Leu Pro
        1245                1250                1255
```

FIG. 1

```
TAC CAT TTA TCA AAG AGA GGG AGA GTG TTA ATG CTT GAG CCT ACC AGA      3966
Tyr His Leu Ser Lys Arg Gly Arg Val Leu Met Leu Glu Pro Thr Arg
    1260                1265                1270

CCA CTC ACA GAT AAC ATG CAC AAG CAA CTG AGA AGT GAA CCA TTT AAC      4014
Pro Leu Thr Asp Asn Met His Lys Gln Leu Arg Ser Glu Pro Phe Asn
1275                1280                1285                1290

TGC TTC CCA ACT TTG AGG ATG AGA GGG AAG TCA ACT TTT GGG TCA TCA      4062
Cys Phe Pro Thr Leu Arg Met Arg Gly Lys Ser Thr Phe Gly Ser Ser
                1295                1300                1305

CCG ATC ACA GTC ATG ACT AGT GGA TTC GCT TTA CAC CAC TTT GCA CGA      4110
Pro Ile Thr Val Met Thr Ser Gly Phe Ala Leu His His Phe Ala Arg
            1310                1315                1320

AAC ATA GCT GAG GTA AAA ACA TAC GAT TTT GTC ATA ATT GAT GAA TGT      4158
Asn Ile Ala Glu Val Lys Thr Tyr Asp Phe Val Ile Ile Asp Glu Cys
        1325                1330                1335

CAT GTG AAT GAT GCT TCT GCT ATA GCG TTT AGG AAT CTA CTG TTT GAA      4206
His Val Asn Asp Ala Ser Ala Ile Ala Phe Arg Asn Leu Leu Phe Glu
    1340                1345                1350

CAT GAA TTT GAA GGA AAA GTC CTC AAA GTG TCA GCC ACA CCA CCA GGT      4254
His Glu Phe Glu Gly Lys Val Leu Lys Val Ser Ala Thr Pro Pro Gly
1355                1360                1365                1370

AGA GAA GTT GAA TTT ACA ACT CAG TTT CCC GTG AAA CTC AAG ATA GAA      4302
Arg Glu Val Glu Phe Thr Thr Gln Phe Pro Val Lys Leu Lys Ile Glu
                1375                1380                1385

GAG GCT CTT AGC TTT CAG GAA TTT GTA AGT TTA CAA GGG ACA GGT GCC      4350
Glu Ala Leu Ser Phe Gln Glu Phe Val Ser Leu Gln Gly Thr Gly Ala
            1390                1395                1400

AAC GCC GAT GTG ATT AGT TGT GGC GAC AAC ATA CTA GTA TAT GTT GCT      4398
Asn Ala Asp Val Ile Ser Cys Gly Asp Asn Ile Leu Val Tyr Val Ala
        1405                1410                1415

AGC TAC AAT GAT GTT GAT AGT CTT GGC AAG CTC CTT GTG CAA AAG GGA      4446
Ser Tyr Asn Asp Val Asp Ser Leu Gly Lys Leu Leu Val Gln Lys Gly
    1420                1425                1430

TAC AAA GTG TCG AAG ATT GAT GGA AGA ACA ATG AAG AGT GGA GGA ACT      4494
Tyr Lys Val Ser Lys Ile Asp Gly Arg Thr Met Lys Ser Gly Gly Thr
1435                1440                1445                1450

GAA ATA ATC ACT GAA GGT ACT TCA GTG AAA AAG CAT TTC ATA GTC GCA      4542
Glu Ile Ile Thr Glu Gly Thr Ser Val Lys Lys His Phe Ile Val Ala
                1455                1460                1465

ACT AAC ATT ATT GAG AAT GGT GTA ACC ATT GAC ATT GAT GTA GTT GTG      4590
Thr Asn Ile Ile Glu Asn Gly Val Thr Ile Asp Ile Asp Val Val Val
            1470                1475                1480

GAT TTT GGG ACT AAG GTT GTA CCA GTT TTG GAT GTG GAC AAT AGA GCG      4638
Asp Phe Gly Thr Lys Val Val Pro Val Leu Asp Val Asp Asn Arg Ala
        1481                1490                1495

GTG CAG TAC AAC AAA ACT GTG GTG AGT TAT GGG GAG CGC ATC CAA AAA      4686
Val Gln Tyr Asn Lys Thr Val Val Ser Tyr Gly Glu Arg Ile Gln Lys
    1500                1505                1510
```

FIG. 1

```
CTC GGT AGA GTT GGG CGA CAC AAG GAA GGA GTA GCA CTT CGA ATT GGC      4734
Leu Gly Arg Val Gly Arg His Lys Glu Gly Val Ala Leu Arg Ile Gly
1515                1520                1525                1530

CAA ACA AAT AAA ACA CTG GTT GAA ATT CCA GAA ATG GTT GCC ACT GAA      4782
Gln Thr Asn Lys Thr Leu Val Glu Ile Pro Glu Met Val Ala Thr Glu
                    1535                1540                1545

GCT GCC TTT CTA TGC TTC ATG TAC AAT TTG CCA GTG ACA ACA CAG AGT      4830
Ala Ala Phe Leu Cys Phe Met Tyr Asn Leu Pro Val Thr Thr Gln Ser
            1550                1555                1560

GTT TCA ACC ACA CTG CTG GAA AAT GCC ACA TTA TTA CAA GCT AGA ACT      4878
Val Ser Thr Thr Leu Leu Glu Asn Ala Thr Leu Leu Gln Ala Arg Thr
        1565                1570                1575

ATG GCA CAG TTT GAG CTA TCA TAT TTT TAC ACA ATT AAT TTT GTG CGA      4926
Met Ala Gln Phe Glu Leu Ser Tyr Phe Tyr Thr Ile Asn Phe Val Arg
    1580                1585                1590

TTT GAT GGT AGT ATG CAT CCA GTC ATA CAT GAC AAG CTG AAG CGC TTT      4974
Phe Asp Gly Ser Met His Pro Val Ile His Asp Lys Leu Lys Arg Phe
1595                1600                1605                1610

AAG CTA CAC ACT TGT GAG ACA TTC CTC AAT AAG TTG GCG ATC CCA AAT      5022
Lys Leu His Thr Cys Glu Thr Phe Leu Asn Lys Leu Ala Ile Pro Asn
                    1615                1620                1625

AAA GGC TTA TCC TCT TGG CTT ACG AGT GGA GAG TAT AAG CGA CTT GGT      5070
Lys Gly Leu Ser Ser Trp Leu Thr Ser Gly Glu Tyr Lys Arg Leu Gly
            1630                1635                1640

TAC ATA GCA GAG GAT GCT GGC ATA AGA ATC CCA TTC GTG TGC AAA GAA      5118
Tyr Ile Ala Glu Asp Ala Gly Ile Arg Ile Pro Phe Val Cys Lys Glu
        1645                1650                1655

ATT CCA GAC TCC TTG CAT GAG GAA ATT TGG CAC ATT GTA GTC GCC CAT      5166
Ile Pro Asp Ser Leu His Glu Glu Ile Trp His Ile Val Val Ala His
    1660                1665                1670

AAA GGT GAC TCG GGT ATT GGG AGG CTC ACT AGC GTA CAG GCA GCA AAG      5214
Lys Gly Asp Ser Gly Ile Gly Arg Leu Thr Ser Val Gln Ala Ala Lys
1675                1680                1685                1690

GTT GTT TAT ACT CTG CAA ACG GAT GTG CAC TCA ATT GCG AGG ACT CTA      5262
Val Val Tyr Thr Leu Gln Thr Asp Val His Ser Ile Ala Arg Thr Leu
                    1695                1700                1705

GCA TGC ATC AAT AGA CGC ATA GCA GAT GAA CAA ATG AAG CAG AGT CAT      5310
Ala Cys Ile Asn Arg Arg Ile Ala Asp Glu Gln Met Lys Gln Ser His
            1710                1715                1720

TTT GAA GCC GCA ACT GGG AGA GCA TTT TCC TTC ACA AAT TAC TCA ATA      5358
Phe Glu Ala Ala Thr Gly Arg Ala Phe Ser Phe Thr Asn Tyr Ser Ile
        1725                1730                1735

CAA AGC ATA TTT GAC ACG CTG AAA GCA AAT TAT GCT ACA AAG CAT ACG      5406
Gln Ser Ile Phe Asp Thr Leu Lys Ala Asn Tyr Ala Thr Lys His Thr
    1740                1745                1750

AAA GAA AAT ATT GCA GTG CTT CAG CAG GCA AAA GAT CAA TTG CTA GAG      5454
Lys Glu Asn Ile Ala Val Leu Gln Gln Ala Lys Asp Gln Leu Leu Glu
1755                1760                1765                1770
```

FIG. 1

```
TTT TCG AAC CTA GCA AAG GAT CAA GAT GTC ACG GGT ATC ATC CAA GAC      5502
Phe Ser Asn Leu Ala Lys Asp Gln Asp Val Thr Gly Ile Ile Gln Asp
                1775                1780                1785

TTC AAT CAC CTG GAA ACT ATC TAT CTC CAA TCA GAT AGC GAA GTG GCT      5550
Phe Asn His Leu Glu Thr Ile Tyr Leu Gln Ser Asp Ser Glu Val Ala
          1790                1795                1800

AAG CAT CTG AAG CTT AAA AGT CAC TGG AAT AAA AGC CAA ATC ACT AGG      5598
Lys His Leu Lys Leu Lys Ser His Trp Asn Lys Ser Gln Ile Thr Arg
     1805                1810                1815

GAC ATC ATA ATA GCT TTG TCT GTG TTA ATT GGT GGT GGA TGG ATG CTT      5646
Asp Ile Ile Ile Ala Leu Ser Val Leu Ile Gly Gly Gly Trp Met Leu
1820                1825                1830

GCA ACG TAC TTC AAG GAC AAG TTC AAT GAA CCA GTC TAT TTC CAA GGG      5694
Ala Thr Tyr Phe Lys Asp Lys Phe Asn Glu Pro Val Tyr Phe Gln Gly
1835                1840                1845                1850

AAG AAG AAT CAG AAG CAC AAG CTT AAG ATG AGA GAG GCG CGT GGG GCT      5742
Lys Lys Asn Gln Lys His Lys Leu Lys Met Arg Glu Ala Arg Gly Ala
                1855                1860                1865

AGA GGG CAA TAT GAG GTT GCA GCG GAG CCA GAG GCG CTA GAA CAT TAC      5790
Arg Gly Gln Tyr Glu Val Ala Ala Glu Pro Glu Ala Leu Glu His Tyr
          1870                1875                1880

TTT GGA AGC GCA TAT AAT AAC AAA GGA AAG CGC AAG GGC ACC ACG AGA      5838
Phe Gly Ser Ala Tyr Asn Asn Lys Gly Lys Arg Lys Gly Thr Thr Arg
     1885                1890                1895

GGA ATG GGT GCA AAG TCT CGG AAA TTC ATA AAC ATG TAT GGG TTT GAT      5886
Gly Met Gly Ala Lys Ser Arg Lys Phe Ile Asn Met Tyr Gly Phe Asp
1900                1905                1910

CCA ACT GAT TTT TCA TAC ATT AGG TTT GTG GAT CCA TTG ACA GGT CAC      5934
Pro Thr Asp Phe Ser Tyr Ile Arg Phe Val Asp Pro Leu Thr Gly His
1915                1920                1925                1930

ACT ATT GAT GAG TCC ACA AAC GCA CCT ATT GAT TTA GTG CAG CAT GAG      5982
Thr Ile Asp Glu Ser Thr Asn Ala Pro Ile Asp Leu Val Gln His Glu
                1935                1940                1945

TTT GGA AAG GTT AGA ACA CGC ATG TTA ATT GAC GAT GAG ATA GAG CCT      6030
Phe Gly Lys Val Arg Thr Arg Met Leu Ile Asp Asp Glu Ile Glu Pro
          1950                1955                1960

CAA AGT CTT AGC ACC CAC ACC ACA ATC CAT GCT TAT TTG GTG AAT AGT      6078
Gln Ser Leu Ser Thr His Thr Thr Ile His Ala Tyr Leu Val Asn Ser
     1965                1970                1975

GGC ACG AAG AAA GTT CTT AAG GTT GAT TTA ACA CCA CAC TCG TCG CTA      6126
Gly Thr Lys Lys Val Leu Lys Val Asp Leu Thr Pro His Ser Ser Leu
1980                1985                1990

CGT GCG AGT GAG AAA TCA ACA GCA ATA ATG GGA TTT CCT GAA AGG GAG      6174
Arg Ala Ser Glu Lys Ser Thr Ala Ile Met Gly Phe Pro Glu Arg Glu
1995                2000                2005                2010

AAT GAA TTG CGT CAA ACC GGC ATG GCA GTG CCA GTG GCT TAT GAT CAA      6222
Asn Glu Leu Arg Gln Thr Gly Met Ala Val Pro Val Ala Tyr Asp Gln
                2015                2020                2025
```

FIG. 1

```
TTG CCA CCA AAG AAT GAG GAC TTG ACG TTT GAA GGA GAA AGC TTG TTT      6270
Leu Pro Pro Lys Asn Glu Asp Leu Thr Phe Glu Gly Glu Ser Leu Phe
            2030                2035                2040

AAG GGA CCA CGT GAT TAC AAC CCG ATA TCG AGC ACC ATT TGT CAT TTG      6318
Lys Gly Pro Arg Asp Tyr Asn Pro Ile Ser Ser Thr Ile Cys His Leu
        2045                2050                2055

ACG AAT GAA TCT GAT GGG CAC ACA ACA TCG TTG TAT GGT ATT GGA TTT      6366
Thr Asn Glu Ser Asp Gly His Thr Thr Ser Leu Tyr Gly Ile Gly Phe
    2060                2065                2070

GGT CCC TTC ATC ATT ACA AAC AAG CAC TTG TTT AGA AGA AAT AAT GGA      6414
Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe Arg Arg Asn Asn Gly
2075                2080                2085                2090

ACA CTG TTG GTC CAA TCA CTA CAT GGT GTA TTC AAG GTC AAG AAC ACC      6462
Thr Leu Leu Val Gln Ser Leu His Gly Val Phe Lys Val Lys Asn Thr
                2095                2100                2105

ACG ACT TTG CAA CAA CAC CTC ATT GAT GGG AGG GAC ATG ATA ATT ATT      6510
Thr Thr Leu Gln Gln His Leu Ile Asp Gly Arg Asp Met Ile Ile Ile
            2110                2115                2120

CGC ATG CCT AAG GAT TTC CCA CCA TTT CCT CAA AAG CTG AAA TTT AGA      6558
Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln Lys Leu Lys Phe Arg
        2125                2130                2135

GAG CCA CAA AGG GAA GAG CGC ATA TGT CTT GTG ACA ACC AAC TTC CAA      6606
Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val Thr Thr Asn Phe Gln
    2140                2145                2150

ACT AAG AGC ATG TCT AGC ATG GTG TCA GAC ACT AGT TGC ACA TTC CCT      6654
Thr Lys Ser Met Ser Ser Met Val Ser Asp Thr Ser Cys Thr Phe Pro
2155                2160                2165                2170

TCA TCT GAT GGC ATA TTC TGG AAG CAT TGG ATT CAA ACC AAG GAT GGG      6702
Ser Ser Asp Gly Ile Phe Trp Lys His Trp Ile Gln Thr Lys Asp Gly
                2175                2180                2185

CAG TGT GGC AGT CCA TTA GTA TCA ACT AGA GAT GGG TTC ATT GTT GGT      6750
Gln Cys Gly Ser Pro Leu Val Ser Thr Arg Asp Gly Phe Ile Val Gly
            2190                2195                2200

ATA CAC TCA GCA TCG AAT TTC ACC AAC ACA AAC AAT TAT TTC ACA AGC      6798
Ile His Ser Ala Ser Asn Phe Thr Asn Thr Asn Asn Tyr Phe Thr Ser
        2205                2210                2215

GTG CCG AAA AAC TTC ATG GAA TTG TTG ACA AAT CAG GAG GCG CAG CAG      6846
Val Pro Lys Asn Phe Met Glu Leu Leu Thr Asn Gln Glu Ala Gln Gln
    2220                2225                2230

TGG GTT AGT GGT TGG CGA TTA AAT GCT GAC TCA GTA TTG TGG GGG GGC      6894
Trp Val Ser Gly Trp Arg Leu Asn Ala Asp Ser Val Leu Trp Gly Gly
2235                2240                2245                2250

CAT AAA GTT TTC ATG AGC AAA CCT GAA GAG CCT TTT CAG CCA GTT AAG      6942
His Lys Val Phe Met Ser Lys Pro Glu Glu Pro Phe Gln Pro Val Lys
                2255                2260                2265

GAA GCG ACT CAA CTC ATG AAT GAA TTG GTG TAC TCG CAA GGG GAG AAG      6990
Glu Ala Thr Gln Leu Met Asn Glu Leu Val Tyr Ser Gln Gly Glu Lys
            2270                2275                2280
```

FIG. 1

```
AGG AAA TGG GTC GTG GAA GCA CTG TCA GGG AAC TTG AGG CCA GTG GCT     7038
Arg Lys Trp Val Val Glu Ala Leu Ser Gly Asn Leu Arg Pro Val Ala
        2285                2290                2295

GAG TGT CCC AGT CAG TTA GTC ACA AAG CAT GTG GTT AAA GGA AAG TGT     7086
Glu Cys Pro Ser Gln Leu Val Thr Lys His Val Val Lys Gly Lys Cys
    2300                2305                2310

CCC CTC TTT GAG CTC TAC TTG CAG TTG AAT CCA GAA AAG GAA GCA TAT     7134
Pro Leu Phe Glu Leu Tyr Leu Gln Leu Asn Pro Glu Lys Glu Ala Tyr
2315                2320                2325                2330

TTT AAA CCG ATG ATG GGA GCA TAT AAG CCA AGT CGA CTT AAT AGA GAG     7182
Phe Lys Pro Met Met Gly Ala Tyr Lys Pro Ser Arg Leu Asn Arg Glu
                2335                2340                2345

GCG TTC CTC AAG GAC ATT CTA AAA TAT GCT AGT GAA ATT GAG ATT GGG     7230
Ala Phe Leu Lys Asp Ile Leu Lys Tyr Ala Ser Glu Ile Glu Ile Gly
            2350                2355                2360

AAT GTG GAT TGT GAC TTG CTG GAG CTT GCA ATA AGC ATG CTC GTC ACA     7278
Asn Val Asp Cys Asp Leu Leu Glu Leu Ala Ile Ser Met Leu Val Thr
        2365                2370                2375

AAG CTC AAG GCG TTA GGA TTC CCA ACT GTG AAC TAC ATC ACT GAC CCA     7326
Lys Leu Lys Ala Leu Gly Phe Pro Thr Val Asn Tyr Ile Thr Asp Pro
    2380                2385                2390

GAG GAA ATT TTT AGT GCA TTG AAT ATG AAA GCA GCT ATG GGA GCA CTA     7374
Glu Glu Ile Phe Ser Ala Leu Asn Met Lys Ala Ala Met Gly Ala Leu
2395                2400                2405                2410

TAC AAA GGC AAG AAG AAA GAA GCT CTC AGC GAG CTC ACA CTA GAT GAG     7422
Tyr Lys Gly Lys Lys Lys Glu Ala Leu Ser Glu Leu Thr Leu Asp Glu
                2415                2420                2425

CAG GAG GCA ATG CTC AAA GCA AGT TGC CTG CGA CTG TAT ACG GGA AAG     7470
Gln Glu Ala Met Leu Lys Ala Ser Cys Leu Arg Leu Tyr Thr Gly Lys
            2430                2435                2440

TTG GGA ATT TGG AAT GGC TCA TTG AAA GCA GAG TTG CGT CCA ATT GAG     7518
Leu Gly Ile Trp Asn Gly Ser Leu Lys Ala Glu Leu Arg Pro Ile Glu
        2445                2450                2455

AAG GTT GAA AAC AAC AAA ACG CGA ACT TTC ACA GCA GCA CCA ATA GAC     7566
Lys Val Glu Asn Asn Lys Thr Arg Thr Phe Thr Ala Ala Pro Ile Asp
    2460                2465                2470

ACT CTT CTT GCT GGT AAA GTT TGC GTG GAT GAT TTC AAC AAT CAA TTT     7614
Thr Leu Leu Ala Gly Lys Val Cys Val Asp Asp Phe Asn Asn Gln Phe
2475                2480                2485                2490

TAT GAT CTC AAC ATA AAG GCA CCA TGG ACA GTT GGT ATG ACT AAG TTT     7662
Tyr Asp Leu Asn Ile Lys Ala Pro Trp Thr Val Gly Met Thr Lys Phe
                2495                2500                2505

TAT CAG GGG TGG AAT GAA TTG ATG GAG GCT TTA CCA AGT GGG TGG GTG     7710
Tyr Gln Gly Trp Asn Glu Leu Met Glu Ala Leu Pro Ser Gly Trp Val
            2510                2515                2520

TAT TGT GAC GCT GAT GGT TCG CAA TTC GAC AGT TCC TTG ACT CCA TTC     7758
Tyr Cys Asp Ala Asp Gly Ser Gln Phe Asp Ser Ser Leu Thr Pro Phe
        2525                2530                2535
```

FIG. 1

```
CTC ATT AAT GCT GTA TTG AAA GTG CGA CTT GCC TTC ATG GAG GAA TGG      7806
Leu Ile Asn Ala Val Leu Lys Val Arg Leu Ala Phe Met Glu Glu Trp
    2540                2545                2550

GAT ATT GGT GAG CAA ATG CTG CGA AAT TTG TAC ACT GAG ATA GTG TAT      7854
Asp Ile Gly Glu Gln Met Leu Arg Asn Leu Tyr Thr Glu Ile Val Tyr
2555                2560                2565                2570

ACA CCA ATC CTC ACA CCG GAT GGT ACT ATC ATT AAG AAG CAT AAA GGC      7902
Thr Pro Ile Leu Thr Pro Asp Gly Thr Ile Ile Lys Lys His Lys Gly
                2575                2580                2585

AAC AAT AGC GGG CAA CCT TCA ACA GTG GTG GAC AAC ACA CTC ATG GTC      7950
Asn Asn Ser Gly Gln Pro Ser Thr Val Val Asp Asn Thr Leu Met Val
            2590                2595                2600

ATT ATT GCA ATG TTA TAC ACA TGT GAG AAG TGT GGA ATC AAC AAG GAA      7998
Ile Ile Ala Met Leu Tyr Thr Cys Glu Lys Cys Gly Ile Asn Lys Glu
        2605                2610                2615

GAG ATT GTG TAT TAC GTC AAT GGC GAT GAC CTA TTG ATT GCC ATT CAC      8046
Glu Ile Val Tyr Tyr Val Asn Gly Asp Asp Leu Leu Ile Ala Ile His
    2620                2625                2630

CCA GAT AAA GCT GAG AGG TTG AGT AGA TTC AAA GAA TCT TTC GGA GAG      8094
Pro Asp Lys Ala Glu Arg Leu Ser Arg Phe Lys Glu Ser Phe Gly Glu
2635                2640                2645                2650

TTG GGC CTG AAA TAT GAA TTT GAC TGT ACC ACC AGG GAC AAG ACA CAG      8142
Leu Gly Leu Lys Tyr Glu Phe Asp Cys Thr Thr Arg Asp Lys Thr Gln
                2655                2660                2665

TTG TGG TTC ATG TCA CAC AGG GCT TTG GAG AGG GAT GGC ATG TAT ATA      8190
Leu Trp Phe Met Ser His Arg Ala Leu Glu Arg Asp Gly Met Tyr Ile
            2670                2675                2680

CCA AAG CTA GAA GAA GAA AGG ATT GTT TCT ATT TTG GAA TGG GAC AGA      8238
Pro Lys Leu Glu Glu Glu Arg Ile Val Ser Ile Leu Glu Trp Asp Arg
        2685                2690                2695

TCC AAA GAG CCG TCA CAT AGG CTT GAA GCC ATC TGT GCA TCA ATG ATT      8286
Ser Lys Glu Pro Ser His Arg Leu Glu Ala Ile Cys Ala Ser Met Ile
    2700                2705                2710

GAA GCA TGG GGT TAT GAC AAG CTG GTT GAA GAA ATC CGC AAT TTC TAT      8334
Glu Ala Trp Gly Tyr Asp Lys Leu Val Glu Glu Ile Arg Asn Phe Tyr
2715                2720                2725                2730

GCA TGG GTT TTG GAA CAA GCG CCG TAT TCA CAG CTT GCA GAA GAA GGA      8382
Ala Trp Val Leu Glu Gln Ala Pro Tyr Ser Gln Leu Ala Glu Glu Gly
                2735                2740                2745

AAG GCG CCA TAT CTG GCT GAG ACT GCG CTT AAG TTT TTG TAC ACA TCT      8430
Lys Ala Pro Tyr Leu Ala Glu Thr Ala Leu Lys Phe Leu Tyr Thr Ser
            2750                2755                2760

CAG CAC GGA ACA AAC TCT GAG ATA GAA GAG TAT TTA AAA GTG TTG TAT      8478
Gln His Gly Thr Asn Ser Glu Ile Glu Glu Tyr Leu Lys Val Leu Tyr
        2765                2770                2775

GAT TAC GAT ATT CCA ACG ACT GAG AAT CTT TAT TTT CAG AGT GGC ACT      8526
Asp Tyr Asp Ile Pro Thr Thr Glu Asn Leu Tyr Phe Gln Ser Gly Thr
    2780                2785                2790
```

FIG. 1

```
GTG GAT GCT GGT GCT GAC GCT GGT AAG AAG AAA GAT CAA AAG GAT GAT      8574
Val Asp Ala Gly Ala Asp Ala Gly Lys Lys Lys Asp Gln Lys Asp Asp
2795                2800                2805                2810

AAA GTC GCT GAG CAG GCT TCA AAG GAT AGG GAT GTT AAT GCT GGA ACT      8622
Lys Val Ala Glu Gln Ala Ser Lys Asp Arg Asp Val Asn Ala Gly Thr
                2815                2820                2825

TCA GGA ACA TTC TCA GTT CCA CGA ATA AAT GCT ATG GCC ACA AAA CTT      8670
Ser Gly Thr Phe Ser Val Pro Arg Ile Asn Ala Met Ala Thr Lys Leu
            2830                2835                2840

CAA TAT CCA AGG ATG AGG GGA GAG GTG GTT GTA AAC TTG AAT CAC CTT      8718
Gln Tyr Pro Arg Met Arg Gly Glu Val Val Val Asn Leu Asn His Leu
        2845                2850                2855

TTA GGA TAC AAG CCA CAG CAA ATT GAT TTG TCA AAT GCT CGA GCC ACA      8766
Leu Gly Tyr Lys Pro Gln Gln Ile Asp Leu Ser Asn Ala Arg Ala Thr
    2860                2865                2870

CAT GAG CAG TTT GCC GCG TGG CAT CAG GCA GTG ATG ACA GCC TAT GGA      8814
His Glu Gln Phe Ala Ala Trp His Gln Ala Val Met Thr Ala Tyr Gly
2875                2880                2885                2890

GTG AAT GAA GAG CAA ATG AAA ATA TTG CTA AAT GGA TTT ATG GTG TGG      8862
Val Asn Glu Glu Gln Met Lys Ile Leu Leu Asn Gly Phe Met Val Trp
                2895                2900                2905

TGC ATA GAA AAT GGG ACT TCC CCA AAT TTG AAC GGA ACT TGG GTT ATG      8910
Cys Ile Glu Asn Gly Thr Ser Pro Asn Leu Asn Gly Thr Trp Val Met
            2910                2915                2920

ATG GAT GGT GAG GAT CAA GTT TCA TAC CCG CTG AAA CCA ATG GTT GAA      8958
Met Asp Gly Glu Asp Gln Val Ser Tyr Pro Leu Lys Pro Met Val Glu
        2925                2930                2935

AAC GCG CAG CCA ACA CTG AGG CAA ATT ATG ACA CAC TTC AGT GAC CTG      9006
Asn Ala Gln Pro Thr Leu Arg Gln Ile Met Thr His Phe Ser Asp Leu
    2940                2945                2950

GCT GAA GCG TAT ATT GAG ATG AGG AAT AGG GAG CGA CCA TAC ATG CCT      9054
Ala Glu Ala Tyr Ile Glu Met Arg Asn Arg Glu Arg Pro Tyr Met Pro
2955                2960                2965                2970

AGG TAT GGT CTA CAG AGA AAC ATT ACA GAC ATG AGT TTG TCA CGC TAT      9102
Arg Tyr Gly Leu Gln Arg Asn Ile Thr Asp Met Ser Leu Ser Arg Tyr
                2975                2980                2985

GCG TTC GAC TTC TAT GAG CTA ACT TCA AAA ACA CCT GTT AGA GCG AGG      9150
Ala Phe Asp Phe Tyr Glu Leu Thr Ser Lys Thr Pro Val Arg Ala Arg
            2990                2995                3000

GAG GCG CAT ATG CAA ATG AAA GCT GCT GCA GTA CGA AAC AGT GGA ACT      9198
Glu Ala His Met Gln Met Lys Ala Ala Ala Val Arg Asn Ser Gly Thr
        3005                3010                3015

AGG TTA TTT GGT CTT GAT GGC AAC GTG GGT ACT GCA GAG GAA GAC ACT      9246
Arg Leu Phe Gly Leu Asp Gly Asn Val Gly Thr Ala Glu Glu Asp Thr
    3020                3025                3030

GAA CGG CAC ACA GCG CAC GAT GTG AAC CGT AAC ATG CAC ACA CTA TTA      9294
Glu Arg His Thr Ala His Asp Val Asn Arg Asn Met His Thr Leu Leu
3035                3040                3045                3050
```

FIG. 1

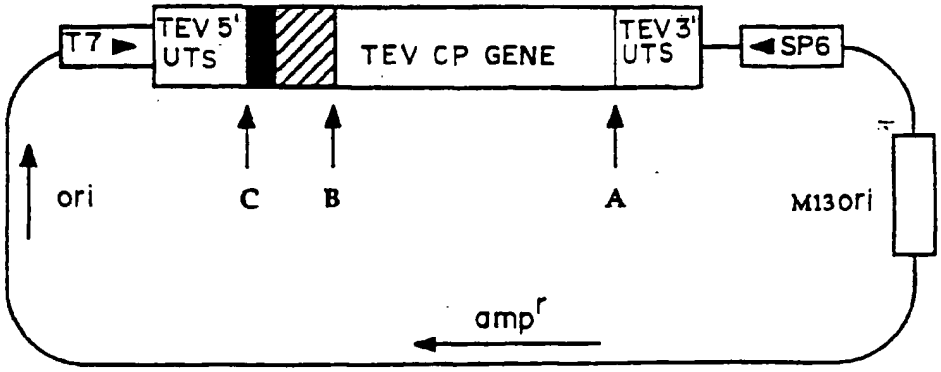
```
GGG GTC CGC CAG TGA TAGTTTCTGC GTGTCTTTGC TTTCCGCTTT TAAGCTTATT      9349
Gly Val Arg Gln

GTAATATATA TGAATAGCTA TTCACAGTGG GACTTGGTCT TGTGTTGAAT AGTATCTTAT    9409

ATATTTTAAT ATGTCTTATT AGTCTCATTA CTTAGGCGAA CGACAAAGTG AGGTCACCTC    9469

GGTCTAATTC TCCTATGTAG TGCGAG                                         9495
```

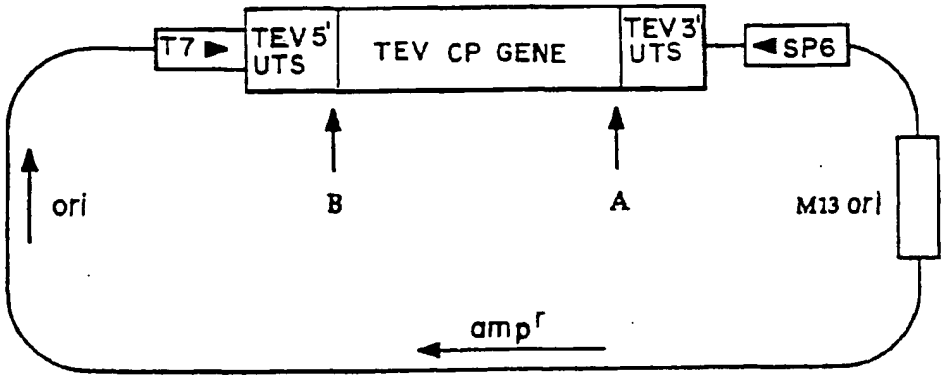
FIG. 1



pTL 37/8595

1. GENERATE BamHI SITE AT A(nt 9312-9317)
2. GENERATE NcoI SITE AT B (nt 8516-8521)
3. GENERATE BamHI SITE (nt 133-138) NcoI SITE (nt 143-148) AND DEOXYADENYLATE RESIDUE (at nt 142) at C.

DIGEST WITH NcoI
REMOVE TEV NUCLEOTIDES 143-200/8462-8516
(FLANKED BY SITES B AND C)AND RELIGATE.



pTC:FL

FIG. 2

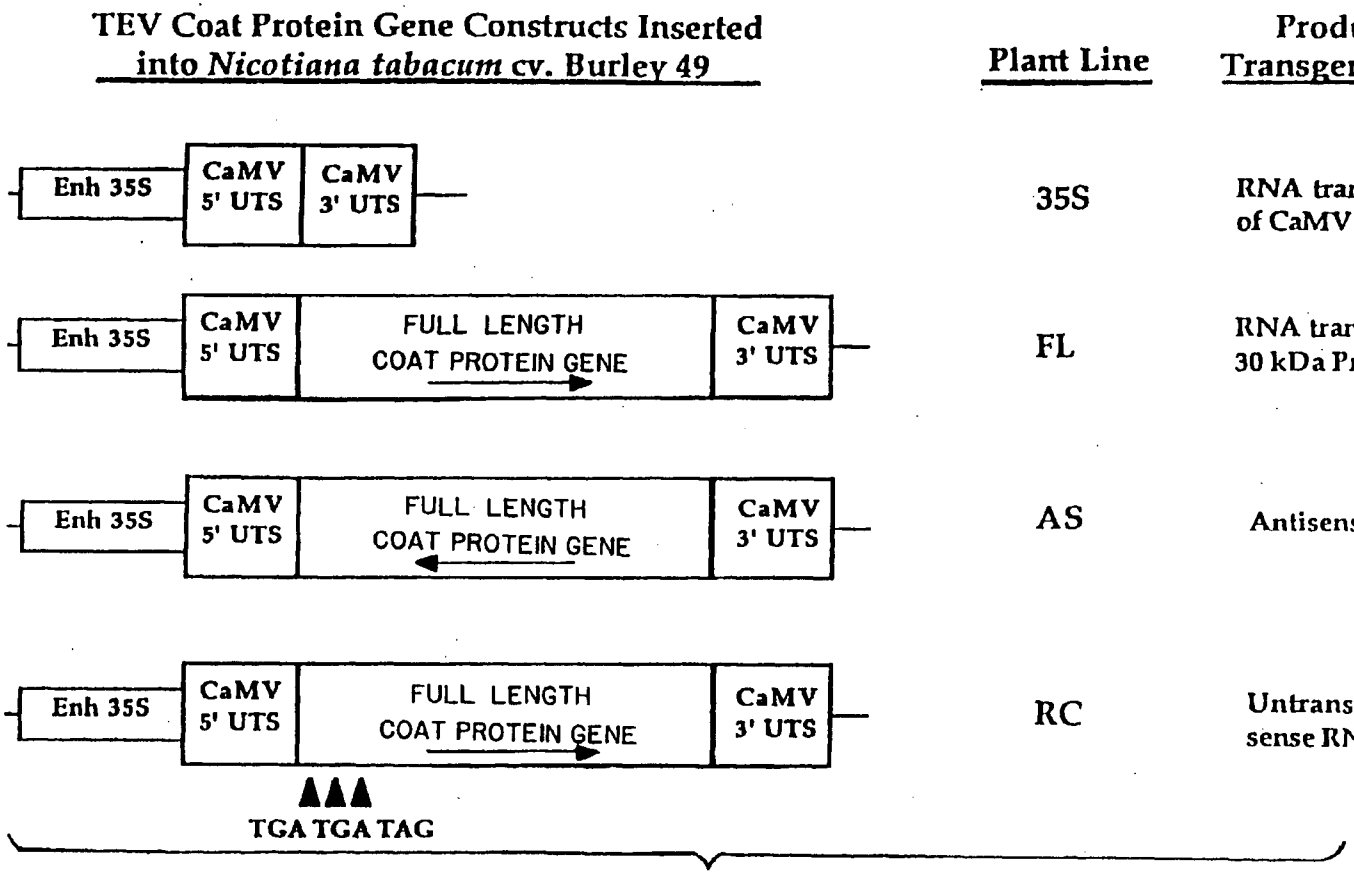

TEV Coat Protein Gene Constructs Inserted into Nicotiana tabacum cv. Burley 49
Plant Line
Product in Transgenic Plants
Enh 35S
CaMV 5' UTS
CaMV 3' UTS
35S
RNA transcript of CaMV UTS
FULL LENGTH COAT PROTEIN GENE
FL
RNA transcript and 30 kDa Protein
AS
Antisense RNA
RC
Untranslatable sense RNA
TGA TGA TAG

FIG. 3

FIG. 4A

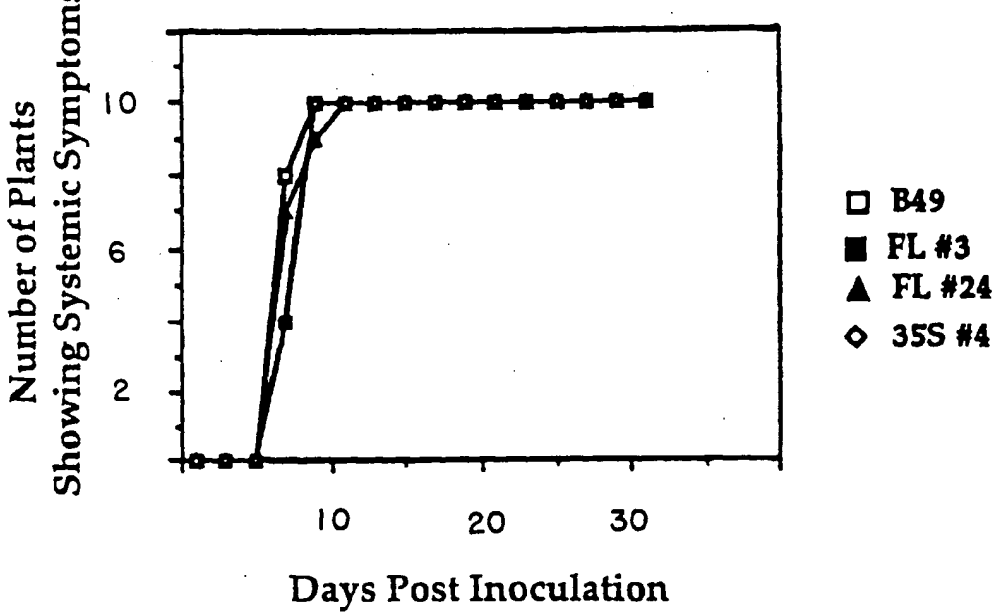
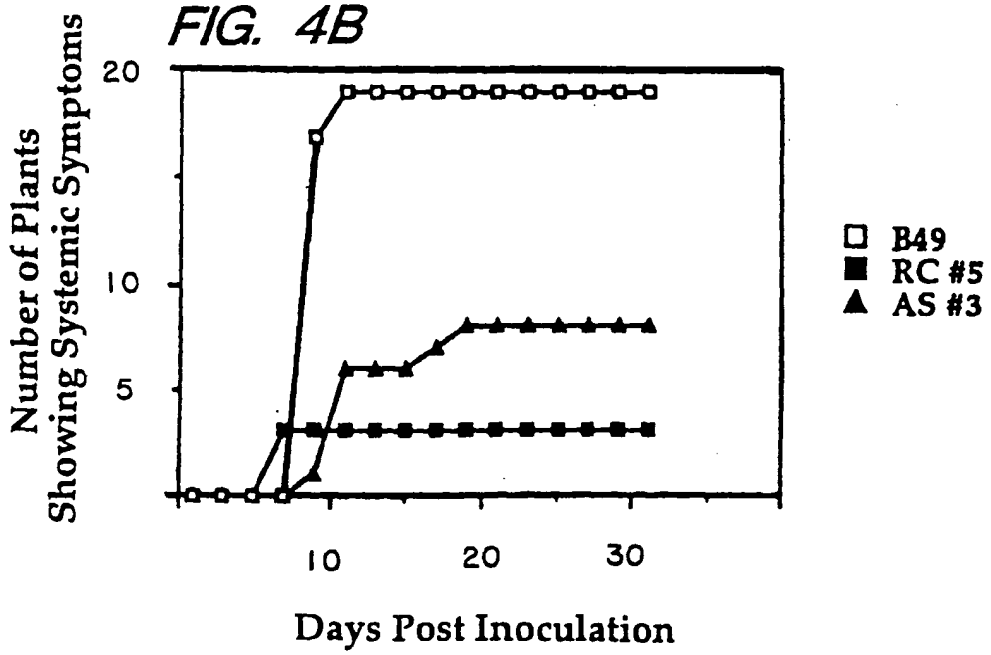

Number of Plants
Showing Systemic Symptoms
10
6
2
10
20
30
Days Post Inoculation
B49
FL #3
FL #24
35S #4

FIG. 4B


Number of Plants
Showing Systemic Symptoms
20
10
5
10
20
30
Days Post Inoculation
B49
RC #5
AS #3